

STIC-Biotech/ChemLib

79093

From: Portner, Ginny  
Sent: Wednesday, October 30, 2002 1:58 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/910,186  
Importance: High

Please search SEQ ID Nos 9 and 10, please back translate SEQ ID NO 10. Thanks.

Ginny Portner  
CM1, Art Unit 1645  
Room 7e13  
Mail box 7e12  
(703) 308-7543

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

RECEIVED  
OCT 30 2002  
STIC

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 11/4/02  
Date Completed: 11/8/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: 141  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 10/6  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpn** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapn** and **.rapn**

*The Pending database search results should not be left in the case because they contain data that is confidential.*

According to the Pre Publication Rules, every patent application received by the United States Patent and Trademark Office after November 29, 2000 will be pre-published at eighteen months from the effective filing date. When the application is published the contents, including the sequences, will become prior art.

Two new databases have been created to hold the pre-published sequences:

**Published\_Applications\_NA** contains nucleic acid sequences; the search results will have the extension **.rnpb**.

**Published\_Applications\_AA** contains amino acid sequences; the search results will have the extension **.rapb**.

Each pre-published application is given a unique Publication Number. An example of a Publication Number is US20021234567A1. The "US" indicates the application was a U.S. application. The first 4 digits show the calendar year the application was published. The next 7 digits represent when the application was published. This 7-digit number starts at zero at the beginning of each calendar year. Each application published is given the next number in order. The "A" indicates a utility patent application and the "1" shows that this was the first time the application had been published. If the applicants submit changes to the application, they may request that the changed application be published again. In such instances, the "1" at the end of the number would be replaced by a "2".

**Sequences in the PGPub database are public information; it is permissible to leave these results in the case.**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OK protein - protein search, using sw model

Run on: November 7, 2002, 13:45:26 ; Search time 65 Seconds  
(without alignments)  
922.504 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MRPNIFSYTNSLLKDI.....NYASLLESTHGWGFPVSE 450

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

908470 seqs, 133250620 residues

Partial number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	2391	100.0	450	21	AA1977136
2	2391	100.0	450	22	AA1977136
3	2386	99.8	1291	20	AA1977136
4	2317	96.9	462	19	AA1977136
5	1899	66.5	399	21	AA1977136
6	867.5	35.3	451	22	AA1977136
7	801.5	33.5	451	21	AA1977136
8	589.5	24.7	449	21	AA1977136
9	589.5	24.7	449	22	AA1977136
10	586.5	24.5	837	21	AA1977136

11	586.5	24.5	1067	21	AA1977136	A manganese supero
12	586.5	24.5	1092	21	AA1977136	A manganese supero
13	586.5	24.5	1296	17	AA1977136	C. botulinum type
14	581	24.3	1070	21	AA1977136	A manganese supero
15	581	24.3	1095	21	AA1977136	A manganese supero
16	581	24.3	1291	19	AA1977136	Clostridium botuli
17	578.5	24.2	452	19	AA1977136	Clostridium botuli
18	577.5	24.2	1295	23	AA1977136	Clostridium botuli
19	575.5	24.1	451	19	AA1977136	Clostridium botuli
20	575.5	24.1	449	21	AA1977136	Synthetic botulinu
21	575.5	24.1	449	22	AA1977136	Botulinum toxin hea
22	575.5	24.1	473	19	AA1977136	Clostridium botuli
23	573	24.0	848	22	AA1977136	Botulinum toxin hea
24	568.5	23.8	847	22	AA1977136	Botulinum toxin hea
25	565	23.6	472	19	AA1977136	Clostridium botuli
26	564	23.6	1059	21	AA1977136	A manganese supero
27	564	23.6	1084	21	AA1977136	A manganese supero
28	558	23.3	1092	22	AA1977136	C. botulinum C2 tr
29	557.5	23.3	432	21	AA1977136	Native botulinum n
30	557.5	23.3	437	22	AA1977136	Botulinum toxin hea
31	557.5	23.3	438	17	AA1977136	Type A neurotoxin
32	557.5	23.3	438	19	AA1977136	Clostridium botuli
33	557.5	23.3	438	21	AA1977136	Synthetic botulinu
34	557.5	23.3	445	19	AA1977136	Clostridium botuli
35	557.5	23.3	462	17	AA1977136	Type A neurotoxin
36	557.5	23.3	462	19	AA1977136	Clostridium botuli
37	556.5	23.3	434	22	AA1977136	Botulinum toxin hea
38	556.5	23.3	435	22	AA1977136	Botulinum toxin hea
39	553.5	23.1	419	22	AA1977136	Botulinum toxin hea
40	551	23.0	1032	22	AA1977136	C. botulinum C2 tr
41	550	23.0	472	19	AA1977136	Clostridium botuli
42	548	22.9	685	22	AA1977136	Modified clostridi
43	546	22.8	431	18	AA1977136	Immunogenic type F
44	546	22.8	432	22	AA1977136	Botulinum toxin hea
45	546	22.8	432	22	AA1977136	Botulinum toxin hea

## ALIGNMENTS

RESULT 1  
ID AA1977136 standard; Protein: 450 AA.  
AC AA1977136;  
XX  
XX  
DT 08-MAY-2000 (first entry)  
XX  
DE Synthetic botulinum neurotoxin serotype C (BoNT/C) C-terminal fragment.  
XX  
KW Botulinum neurotoxin; heavy chain; BoNT; serotype C;  
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
KW VEE; botulinism; vaccine; diagnosis; drug screening.  
XX  
OS Clostridium botulinum.  
OS Synthetic.  
PN WO200002524-A2.  
XX  
PD 20-JAN-2000.  
XX  
PF 09-JUL-1999; 99WO-US15570.  
XX  
PR 10-JUL-1998; 98US-0092416.  
PR 12-MAY-1999; 99US-0133870.  
XX  
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;  
WPI: 2000-160827/14.  
DR N-PSDB; AA287214.  
XX



Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum toxin serotypes A-G, is used for inducing an immune response against botulinum -

Claim 24; Page 41-42; 54pp; English.

The invention relates to novel vaccines that induce a protective immune response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant DNA construct comprising a vector, and at least one nucleic acid fragment comprising a C-terminal heavy chain fragment (HC) from BoNT serotypes A-G. In preferred embodiments of the invention, the vector is a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this vector results in the production of large amounts of a protein encoded by a sequence cloned into the replicon. The constructs are used to produce vaccines against botulinum. The proteins can also be used as diagnostic tools for the diagnosis of botulinum. The transformed host cells can be used to analyse the effectiveness of drugs and agents which inhibit toxin effects. The vaccine currently used against botulinum is dangerous and expensive to produce, and contains formalin, which is very painful for the recipient. Also, the vaccine is incomplete, in that only 5 of the 7 serotypes are represented in the formulation. The novel vaccine of overcomes these problems, as it is easily purified, and available in large quantities. It is also expressed in the lymph nodes for a better immune response. Sequences AAY7134-Y7139 represent synthetic BoNT HC fragments used in the present invention. The DNA encoding these sequences had been optimised for codon usage for expression in yeast.

Sequence 450 AA:

Query Match 100.0%; Score 2391; DB 21; Length 450;  
Best Local Similarity 100.0%; Pred. No. 9.6e-164;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MTIPNIFSYNNLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVEEGDVL 60  
1 MTIPNIFSYNNLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVEEGDVL 60  
61 NPIFFDFKLGSSGDRGKVTQNTQENIVNSMYESFSISFWIRINKWVSNLPGYTIIDS 120  
61 NPIFFDFKLGSSGDRGKVTQNTQENIVNSMYESFSISFWIRINKWVSNLPGYTIIDS 120  
121 VKNNSGWSIGIISNPLVFTLKQNEDESSQSFNSYDINNNAPGYNKWFVFTVNNMGNMK 180  
121 VKNNSGWSIGIISNPLVFTLKQNEDESSQSFNSYDINNNAPGYNKWFVFTVNNMGNMK 180  
181 IYNGKLIDTIKVKELTGINFSTKITEINKIPDTGLITSDSNINNMWIRDFYIFAKELD 240  
181 IYNGKLIDTIKVKELTGINFSTKITEINKIPDTGLITSDSNINNMWIRDFYIFAKELD 240  
241 GKDINLFNSLQYTNVVDYNGDLRYNKEYYMYNIDYLNRYMYANSRQVFNTRNNND 300  
241 GKDINLFNSLQYTNVVDYNGDLRYNKEYYMYNIDYLNRYMYANSRQVFNTRNNND 300  
301 FNEGYKILIKRIGNTNTRVGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIYAI 360  
301 FNEGYKILIKRIGNTNTRVGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIYAI 360  
361 GUREQKIDNDLIFQIOPMNTYYASQIFKSNFNGENTSGICISGYFRLLGGDWYRH 420  
361 GUREQKIDNDLIFQIOPMNTYYASQIFKSNFNGENTSGICISGYFRLLGGDWYRH 420  
421 NVLPTVKOGNVASLLESTSHWGVPVSE 450  
421 NVLPTVKOGNVASLLESTSHWGVPVSE 450

RESULT 2  
AAB04092  
ID AAB04092 standard; Protein; 450 AA.  
XX  
AC AAB04092;

XX 11-APR-2001 (first entry)  
DT Botulinum toxin heavy chain C-terminal sequence (serotype C).  
DE  
XX Botulinum toxin; neurotoxin; heavy chain; recombinant expression;  
KW recombinant vector; antigen; immune response; vaccine; bacterium;  
KW infection.  
XX  
XX Synthetic.  
OS Clostridium botulinum.  
XX WO2000067700-A2.  
XX  
XX 16-NOV-2000.  
XX 12-MAY-2000; 2000WO-US12890.  
XX 12-MAY-1999; 99US-0133865.  
PR 12-MAY-1999; 99US-0133866.  
PR 12-MAY-1999; 99US-0133867.  
PR 12-MAY-1999; 99US-0133868.  
PR 12-MAY-1999; 99US-0133869.  
PR 12-MAY-1999; 99US-0133873.  
PR 29-JUL-1999; 99US-0146192.  
XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
FA Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
XX WPI; 2001-016048/02.  
XX N-PSDB; AAA54486.  
XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
PT vaccine against botulinum  
XX  
PS Claim 3; Fig 5b; 73pp; English.  
XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
CC chain and then posttranslationally nicked, forming a dichain  
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
CC remain linked by a disulfide bond. Nucleic acids encoding the  
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
CC expression vectors and expressed in transformed cells to produce  
CC peptide antigens useful for eliciting an immune response to give  
CC protective immunity against botulinum neurotoxin, which causes  
CC botulinum. The nucleic acids are expressible in a recombinant  
CC organisms such as Escherichia coli or Pichia pastoris. The use  
CC of recombinant nucleic acids are advantageous since it eliminates  
CC the need to culture large quantities of hazardous toxin-producing  
CC bacterium. Production yield from the genetically engineered product  
CC is also high and cost of production is lower. The nucleic acids can  
CC be derived from Clostridium botulinum serotypes A-G.

Sequence 450 AA:

Query Match 100.0%; Score 2391; DB 22; Length 450;  
Best Local Similarity 100.0%; Pred. No. 9.6e-164;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTIPNIFSYNNLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVEEGDVL 60  
DB 1 MTIPNIFSYNNLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEVEEGDVL 60  
QY 61 NPIFFDFKLGSSGDRGKVTQNTQENIVNSMYESFSISFWIRINKWVSNLPGYTIIDS 120  
DB 61 NPIFFDFKLGSSGDRGKVTQNTQENIVNSMYESFSISFWIRINKWVSNLPGYTIIDS 120  
QY 121 VKNNSGWSIGIISNPLVFTLKQNEDESSQSFNSYDINNNAPGYNKWFVFTVNNMGNMK 180  
DB 121 VKNNSGWSIGIISNPLVFTLKQNEDESSQSFNSYDINNNAPGYNKWFVFTVNNMGNMK 180

CC PCR and site-directed mutagenesis. The modified recombinant  
 CC botulinum toxin maintains its ability to translocate from the gut  
 CC into the general circulation but is non-toxic. It can be used as  
 CC an oral vaccine for antigenic peptides including botulinum toxin  
 CC (i.e. an oral vaccine for botulism) or for the oral delivery of  
 CC other therapeutic agents to the general circulation.

XX Sequence 1291 AA;

QY 181 IYINGKLIDTIKVELTGINFSTIPEINKIPDTGLITSDSDININMWIRDFVIFAKELD 240  
 DB 181 IYINGKLIDTIKVELTGINFSTIPEINKIPDTGLITSDSDININMWIRDFVIFAKELD 240  
 QY 241 GKDNILFNSI-QYTNVVKDYWGNDLRNKEVYVYNDYLNRYMYANSROIIVFTRNNND 300  
 DB 241 GKDNILFNSI-QYTNVVKDYWGNDLRNKEVYVYVNDYLNRYMYANSROIIVFTRNNND 300  
 QY 301 FNEGYKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 360  
 DB 301 FNEGYKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 360  
 QY 361 GLREOTKDINDNIIFQIPMNNTYYASQIFKSNFNGENISGICSTGYRFRGGDWYRH 420  
 DB 361 GLREOTKDINDNIIFQIPMNNTYYASQIFKSNFNGENISGICSTGYRFRGGDWYRH 420  
 421 NYLVTVKQGNVYASLLESTSTHWGFVPVSE 450  
 421 NYLVTVKQGNVYASLLESTSTHWGFVPVSE 450  
 RESULT 3  
 AAY05814  
 AAY05814 standard; Protein: 1291 AA.  
 AAY05814;  
 02-AUG-1999 (first entry)  
 Non-toxic modified botulinum toxin rBoNT/C.  
 Botulinum toxin; botulism; rBoNT/C; vaccine; drug delivery;  
 Clostridium botulinum.  
 Synthetic.  
 Key Location/Qualifiers  
 Misc-difference 229 /note= "His in native toxin"  
 Misc-difference 230 /note= "Glu in native toxin"  
 Misc-difference 233 /note= "His in native toxin"

WO9920306-A1.  
 29-APR-1999.  
 16-OCT-1998; 98WO-US21897.  
 20-OCT-1997; 97US-0954302.  
 (UYJE-) UNIV JEFFERSON THOMAS.  
 Kiyatkin N, Maksymowych A, Simpson L;  
 WPI; 1999-302646/25.  
 N-PSDB; AAX25521.  
 Modified toxin useful for systemic delivery of oral vaccines and  
 therapeutic agents  
 Example 1; Page 31-33; 37pp; English.

The present sequence represents a modified serotype C botulinum  
 toxin, termed rBoNT/C, in which amino acids His-229, Glu-230 and  
 His-233 of the native sequence are substituted by Gly, Thr and  
 Asn, respectively, i.e. the zinc binding motif (see AAY05817) of  
 the light chain holotoxin is modified, resulting in loss of  
 endoprotease activity. DNA coding for the modified botulinum toxin  
 (see AAX25521) was assembled from 3 separate toxin fragments using

CC Botulinum toxin maintains its ability to translocate from the gut  
 CC into the general circulation but is non-toxic. It can be used as  
 CC an oral vaccine for antigenic peptides including botulinum toxin  
 CC (i.e. an oral vaccine for botulism) or for the oral delivery of  
 CC other therapeutic agents to the general circulation.

XX Sequence 1291 AA;

QY 2 TIPNITSYNNLSLLKDIINEYFNININDSKILSLQNRKNTLVDTSYNAEVSSEGVDQLN 61  
 DB 843 TIPNITSYNNLSLLKDIINEYFNININDSKILSLQNRKNTLVDTSYNAEVSSEGVDQLN 902  
 QY 62 PIFPFDKLGSSGEDRGKVIIVTQENIVYNSMYESTSISFWIRINKWVSNLPGYTIIDSV 121  
 DB 903 PIFPFDKLGSSGEDRGKVIIVTQENIVYNSMYESTSISFWIRINKWVSNLPGYTIIDSV 962  
 QY 122 KNSGWSIGIISNLFVTLKONEDSEOSINFSYDISNNAFGYNKWFVTVTNMGMNMI 181  
 DB 963 KNSGWSIGIISNLFVTLKONEDSEOSINFSYDISNNAFGYNKWFVTVTNMGMNMI 1022  
 QY 182 YINGKLIDTIKVELTGINFSTIPEINKIPDTGLITSDSDININMWIRDFVIFAKELD 241  
 DB 1023 YINGKLIDTIKVELTGINFSTIPEINKIPDTGLITSDSDININMWIRDFVIFAKELD 1082  
 QY 242 KDINILFNSLOQYTNVVKDYWGNDLRNKEVYVYNDYLNRYMYANSROIIVFTRNNND 301  
 DB 1083 KDINILFNSLOQYTNVVKDYWGNDLRNKEVYVYNDYLNRYMYANSROIIVFTRNNND 1142  
 QY 302 NEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 361  
 DB 1143 NEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1202  
 QY 362 LREOTKDINDNIIFQIPMNNTYYASQIFKSNFNGENISGICSTGYRFRGGDWYRH 421  
 DB 1203 LREOTKDINDNIIFQIPMNNTYYASQIFKSNFNGENISGICSTGYRFRGGDWYRH 1262  
 QY 422 YLVPVTKQGNVYASLLESTSTHWGFVPVSE 450  
 DB 1263 YLVPVTKQGNVYASLLESTSTHWGFVPVSE 1291

RESULT 4  
 AAW68397  
 ID AAW68397 standard; Protein: 462 AA.  
 XX AAW68397;  
 XX AAW68397;  
 DT 07-DEC-1998 (first entry)  
 XX Clostridium botulinum type C1 toxin C fragment.  
 DE Clostridium botulinum type C1 toxin C fragment.  
 KW Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;  
 KW botulism; BotC.  
 XX Clostridium botulinum serotype C1 strain Stockholm.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FH Peptide 1..21  
 FT /note= "N-terminal His tag"  
 XX WO9808540-A1.  
 XX 05-MAR-1998.  
 XX 28-AUG-1997; 97WO-US15394.  
 XX 28-AUG-1996; 96US-0704159.  
 PR

XX PA (OPHI-) OPHIDIAN PHARM INC.  
XX PI Thalley BS, Williams JA;  
XX DR WPI; 1998-230234/20.  
XX DR N-PSDB; AAV30588.  
XX PT Host cell containing recombinant expression vector encoding  
XX PT Clostridium botulinum type B or E toxin - useful to treat humans  
XX PT and other animals at risk of intoxication with clostridial toxin  
XX PS Example 45; Page 339-341; 428pp; English.  
XX CC This is the amino acid sequence of the histidine-tagged C fragment  
XX CC of Clostridium botulinum (Stockholm strain) type C1 neurotoxin,  
XX CC encoded by a DNA sequence (see AAV30588) in plasmid pTH3b. This  
XX CC vector is used to express BcC soluble C fragment in Escherichia  
XX CC coli host cells, and the recombinant C fragment was purified on an  
XX CC affinity column. The invention relates to recombinant proteins  
XX CC derived from C. botulinum toxins, especially type B and type E  
XX CC toxins. Methods are provided which allow for the isolation of  
XX CC soluble recombinant proteins free of significant endotoxin  
XX CC contamination. Preferred hosts for production of recombinant  
XX CC proteins are E. coli, insect cells and yeast cells. The  
XX CC recombinant toxins are used as immunogens for the production of  
XX CC vaccines and antitoxins that are useful in the treatment of humans  
XX CC and animals at risk of intoxication with clostridial toxin.  
XX CC Sequence 462 AA;  
XX CC Query Match 96.98; Score 2317; DB 19; Length 462;  
XX CC Best Local Similarity 99.88; Pred. No. 2e-156;  
XX CC Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
XX CC 14 SLKDIINEYNNINDSKILSLQNRKNTLVDTSYNAEVEEGDVLNLPFPDFKLGS 73  
XX CC 26 ALLKDIINEYNNINDSKILSLQNRKNTLVDTSYNAEVEEGDVLNLPFPDFKLGS 85  
XX CC 74 GDRGKVTITQENIVNYSMTESFSISFWIRINKWNSLPGYIIDSVKNSGWSIGIS 133  
XX CC 86 GDRGKVTITQENIVNYSMTESFSISFWIRINKWNSLPGYIIDSVKNSGWSIGIS 145  
XX CC 134 NFLVFTLKONEDSEGSINFSYDIENNAPGYNKFEFTVTNNMGNKIYINGKLIDTIKV 193  
XX CC 146 NFLVFTLKONEDSEGSINFSYDIENNAPGYNKFEFTVTNNMGNKIYINGKLIDTIKV 205  
XX CC 194 KELTGINFSKTITFEINKIPDTGLITSDSDINNMWIRDFYIFAKELDKDINILFNSLOY 253  
XX CC 206 KELTGINFSKTITFEINKIPDTGLITSDSDINNMWIRDFYIFAKELDKDINILFNSLOY 255  
XX CC 254 TNVVKDYNGNDLRYNKEYMYNIDYLNRYMYANSRQIVENTRRNNDFNKGKIIKRIK 313  
XX CC 266 TNVVKDYNGNDLRYNKEYMYNIDYLNRYMYANSRQIVENTRRNNDFNKGKIIKRIK 325  
XX CC 314 GNTNDRVRGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIATGLREQTKDINDNI 373  
XX CC 326 GNTNDRVRGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIATGLREQTKDINDNI 385  
XX CC 374 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYFRGLGGDWYRNLYLPTVKQNYA 433  
XX CC 386 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYFRGLGGDWYRNLYLPTVKQNYA 445  
XX CC 434 SLESTSTHWGFPVSE 450  
XX CC 446 SLESTSTHWGFPVSE 462  
XX CC RESULT 5  
XX CC ID AAV78982  
XX CC AAV78982 standard; Protein; 399 AA.  
XX AC AAV78982;

XX 20-JUN-2000 (first entry)  
XX DT C. botulinum type D toxin amino acid sequence.  
XX DE Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;  
XX KW protect.  
XX KW Clostridium botulinum.  
XX OS WO200005252-A1.  
XX PN 03-FEB-2000.  
XX PD 20-JUL-1999; 99WO-IB01301.  
XX PF 22-JUL-1998; 98ZA-0006538.  
XX PR (AGRI-) AGRIC RES COUNCIL.  
XX PA De Bruyn EE, Botha AD;  
XX PI WPI; 2000-205375/18.  
XX PT N-PSDB; AA298630, AA298631.  
XX PT Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin  
XX PT type D, useful in vaccines for protection against botulism, comprises  
XX PT at least one amino acid mutation not present in the wild type D  
XX PT neurotoxins -  
XX PS Claim 3, 4; Page 54-57; 66pp; English.  
XX CC This sequence represents the amino acid sequence of a synthetic  
XX CC non-toxic immunogenic derivative of Clostridium botulinum type D toxin  
XX CC (BoNT). Botulinum neurotoxin causes botulism poisoning in cattle and  
XX CC sheep, and usually results in the death of the affected or poisoned  
XX CC animal. The non-toxic immunogenic fragments of the C. botulinum  
XX CC neurotoxin are useful in vaccines to protect animals (e.g. humans,  
XX CC cattle, sheep, pigs) against BoNT type D poisoning. The non-toxic  
XX CC fragments can be produced relatively simply and inexpensively  
XX CC (specifically by fermentation techniques). As the fragments are not  
XX CC toxic the risk to production staff is reduced.  
XX CC Sequence 399 AA;  
XX CC Query Match 66.5%; Score 1589; DB 21; Length 399;  
XX CC Best Local Similarity 76.08; Pred. No. 3.3e-106;  
XX CC Matches 307; Conservative 33; Mismatches 56; Indels 8; Gaps 4;  
XX CC 50 AEVSEEGDVLNLPFPDFKLGSSEDRGKVITQENIVNYSMTESFSISFWIRINKWY 109  
XX CC 1 AEVRVEGNVQLNLPFPDFKLGSSEDRGKVITQENIVNYSMTESFSISFWIRINKWY 60  
XX CC 110 SNLPGYTIIDSVKNSGWSIGIISNLFVTLKONEDSEGSINFSYDIENNAPGYNKFEV 169  
XX CC 61 SNLPGYTIIDSVKNSGWSIGIISNLFVTLKONEDSEGSINFSYDIENNAPGYNKFEV 120  
XX CC 170 TVTNMGMNKIYINGKLIDTIKYKELTGINFSKTITFEINKIPDTGLITSDSDINMWI 229  
XX CC 121 ITTNMGMNKIYINGKLIDTIKYKELTGINFSKTITFEINKIPDTGLITSDSDINMWI 180  
XX CC 230 RQFYIFAKELDKDINILFNSLOYTNVVKDYNGNDLRYNKEYMYNIDYLNRYMYANSRQ 289  
XX CC 181 RQFYIFAKELDKDINILFNSLOYTNVVKDYNGNDLRYNKEYMYNIDYLNRYMYANSRQ 240  
XX CC 290 IVFNTRNNDFNKGKIIKRIKRIKRIKRIKRIKRIKRIKRIKRIKRIKRIKRIKRIK 349  
XX CC 241 IVFNTRNNDFNKGKIIKRIKRIKRIKRIKRIKRIKRIKRIKRIKRIKRIKRIKRIK 295  
XX CC 350 DNHST-EDIYAGLRQTKD-INDNIIFQIQPMNNTYYASQIFKSNFNGENISGICISG 407  
XX CC 296 PSRNLGTDLVPLGALDQPMDEIRKYSFIQPCNTFDYASQLFSLSSNATNRLGILSIG 355

Qy 408 TYRRLGGD-WYRHNLYPTVKOGNYASLLESTSTHWGFVPVSE 450  
 Db 356 SYSEKLGDDYWFNHEYLIPVIKIEHVASLLESTSTHWGFVPVSE 399

RESULT 6  
 AAB04093  
 ID AAB04093 standard; Protein: 451 AA.

XX AAB04093;  
 XX 11-APR-2001 (first entry)  
 XX Botulism toxin heavy chain C-terminal sequence (serotype D).  
 XX Botulism; toxin; neurotoxin; heavy chain; recombinant expression;  
 XX recombinant vector; antigen; immune response; vaccine; bacterium;  
 XX infection.

Synthetic.  
 Clostridium botulinum.  
 WO200067700-A2.  
 16-NOV-2000.  
 12-MAY-2000; 2000WO-US12890.  
 12-MAY-1999; 99US-0133865.  
 12-MAY-1999; 99US-0133866.  
 12-MAY-1999; 99US-0133867.  
 12-MAY-1999; 99US-0133868.  
 12-MAY-1999; 99US-0133869.  
 12-MAY-1999; 99US-0133873.  
 29-JUL-1999; 99US-0146132.

(USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
 Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
 WPI: 2001-016048/02.  
 N-PSDB; AAA54487.  
 New nucleic acids encoding the carboxy- or amino-terminal portions of  
 the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 vaccine against botulism

Claim 3; Fig 6b; 73pp; English.  
 Botulism neurotoxins are translated as a single 150 kDa polypeptide  
 chain and then posttranslationally nicked, forming a dichain  
 consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 remain linked by a disulfide bond. Nucleic acids encoding the  
 carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 expression vectors and expressed in transformed cells to produce  
 peptide antigens useful for eliciting an immune response to give  
 protective immunity against botulinum neurotoxin, which causes  
 botulism. The nucleic acids are expressible in a recombinant  
 organisms such as Escherichia coli or Pichia pastoris. The use  
 of recombinant nucleic acids are advantageous since it eliminates  
 the need to culture large quantities of hazardous toxin-producing  
 bacterium. Production yield from the genetically engineered product  
 is also high and cost of production is lower. The nucleic acids can  
 be derived from Clostridium botulinum serotypes A-G.

Sequence 451 AA;  
 Query Match 36.3%; Score 867.5; DB 22; Length 451;  
 Best Local Similarity 43.2%; Pred. No. 2.5e-54;  
 Matches 195; Conservative 84; Mismatches 145; Indels 27; Gaps 11;  
 Qy 2 TTPNFNYSYNNLLKDIINEYFNNSKILSLQNRKNTLVDTSYNAEYSEGDVQLN 61

Db 14 TPNPFNYSYNNLLKDIINEYFNNSKILSLQNRKNTLVDTSYNAEYSEGDVQLN 73  
 Qy 62 PIFPFDFKLGSGEDRGKIVTQENIVNYSYSEFSISFWIRINKWVSNLPG-YTIIDS 120  
 Db 74 TIYTNDFKLSGGD--KIIVNLLNNILYSAIYENSVSFWIKISKDLTNSHNEVTIINS 130  
 Qy 121 VKNNSGWSIGIISNLFVTLKQNEDESEQSINFSDYDINNAPGY-NKWEFVTVTNMNGNM 179  
 Db 131 IEQNSGKLCIRNGNIEWILODVNRKYSLIFDYSLSLSTGTNTKWEFFVTIINNMGYK 190  
 Qy 180 KIYNGKLIDTIKVKELTGINFKTIIFKINIPDITGLITSDSDNINMIRDFYIFAKEL 239  
 Db 191 KLYINGELKQSKIEDLDEYKDKTIVFGIDENID-----ENQMLWIRDFNIFSKEL 242  
 Qy 240 DKGIDNIFLSQTVNKKVQWGNLDRLYKEYYVWIDILNRYMYANSRQIVENTRRNN 299  
 Db 243 SNEDINIVYEGQILRNVIKDYWGNCPLAFDEYIINDYIDRYIAPESNVLVLVQYPPDRS 302  
 Qy 300 DFNEGYKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFKMKNETMYA---DNHSTED 356  
 Db 303 KLYTGNPTIKSVSDKNPYSKILNGDNILHMLYNSRKYMIIRDITVIATOGGDCSQNC 362  
 Qy 357 IYA:GLREQTKDINDNTIFOIQPMNNTYIYASQIFKSNFNGENISIGCSI-GTYRFLGG 415  
 Db 363 VYALKQSLNLTGNYGIG-IFSKITVSKNKYCSQIF-SSFR-ENTMLLADIYKPFRES--- 416  
 Qy 416 DWYRHNLYPTVKOGNYASLLESTSTHWGFV 446  
 Db 417 --FRNATPVAVT--NYETKLLSTSSFWKFI 443

RESULT 7  
 AAW68398  
 ID AAW68398 standard; Protein: 451 AA.  
 XX AAW68398;  
 XX 07-DEC-1998 (first entry)  
 XX Clostridium botulinum type D toxin C fragment.  
 DE Antitoxin; vaccine; neurotoxin; toxin D; intoxication; immunogen;  
 KW botulism; BoD.  
 XX Clostridium botulinum serotype D.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 XX Peptide 1..21  
 XX /note= "N-terminal His tag"  
 XX WO9808540-A1.  
 XX 05-MAR-1998.  
 XX 28-AUG-1997; 97WO-US15394.  
 XX 28-AUG-1996; 96US-0704159.  
 XX (OPHI-) OPHIDIAN PHARM INC.  
 XX Thalley BS, Williams JA;  
 XX WPI: 1998-230234/20.  
 XX N-PSDB; AAV30591.  
 XX Host cell containing recombinant expression vector encoding  
 XX Clostridium botulinum type B or E toxin - useful to treat humans  
 XX and other animals at risk of intoxication with clostridial toxin  
 XX Example 47; Page 352-353; 428pp; English.  
 XX







Qy 343 KNETMYADNTHSTEDIIAIGUREQTKDINDNITFQIQPMNNTYYASQIFKSNFNGENISG 402  
 Db 965 -----ATNASQAGVERKILSALETPDVGNSLQVVMKSKNDQGITNKC-KMNLQDNNGND 1017  
 Qy 403 ICSIGTYRFRGLGWDYRHNLYPTVYKQGNVYASLESTST-----THMGFVPVSE 450  
 Db 1018 IGFICGPHQF-----NNIAKLVASNWTNROIERSSRTLGCSEWEIFPDD 1060

RESULT 12  
 AAY93310  
 ID AAY93310 standard; protein; 1092 AA.  
 XX  
 AC AAY93310;  
 XX

04-SEP-2000 (first entry)

A manganese superoxide dismutase (Mn-SOD) construct.

Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
 neuronal cell targeting component; NCIC; neuronal disease;  
 oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
 Huntington's disease; motor neurone disease;  
 botulinum neurotoxin serotype A.

Synthetic.  
 Homo sapiens.  
 Bacillus stearothermophilus.  
 Clostridium botulinum.

WO200028041-A1.

18-MAY-2000.

05-NOV-1999; 99WO-GB03699.

05-NOV-1998; 98GB-0024282.

(MICR-) MICROBIOLOGICAL RES AUTHORITY.

Shone CC, Sutton JM, Hallis B, Silman N;

WPI; 2000-376553/32.

Novel composition, comprising superoxide dismutase linked by a  
 cleavable linker to a neuronal cell targeting component useful for  
 delivering superoxide dismutase to neuronal cells to treat ischemia -

Disclosure; Page 51-54; 65pp; English.

The present sequence represents a construct of the invention, comprising  
 a mitochondrial leader sequence from human manganese superoxide  
 dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker  
 that can be cleaved by thrombin, and a heavy chain derived from  
 botulinum neurotoxin serotype A. The specification describes a  
 composition for delivery of SOD to neuronal cells. The composition  
 comprises SOD linked, by a cleavable linker, to a neuronal cell  
 targeting component (NCTC). This component has a domain that binds  
 to a neuronal cell and a domain that translocates the SOD of the  
 composition into the neuronal cell. After translocation, the linker  
 is cleaved to release the SOD. The composition is useful for treating  
 neuronal diseases caused or augmented by oxidative stress, such as  
 ischemic stroke, trauma, Parkinson's disease, Huntington's disease and  
 motor neurone diseases.

Sequence 1092 AA;

Query Match 24.5%; Score 586.5; DB 21; Length 1092;  
 Best Local Similarity 31.6%; Pred. No. 1.1e-33;  
 Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Qy 3 IPFNIFSTNNSLLKDIINEYFNINNSKILSLQNRKNTLVDSGYNAEVSEGDVQLNP 62

Db 645 IPQLSKYVDNQALLSTFTYIKNIINTSILNRYESNHLIDLSRYASKINIGSKVNDP 704  
 Qy 63 IFPPD---FKLSSGEDRGKVIYVTONENIVNSWYSPFSISFWIRINKVWSNLP---GYT 116  
 Db 705 IDKNOICLFNLESS---KIEVILKNAIVNSMYENFSTFWIRIPKYPFNSISLNEYT 759  
 Qy 117 IIDSVKNSGWSIGIISNLFVFLTKQNEDESEQSINFSYDISNAPGY-NKWEFFVTVNNM 175  
 Db 760 IINCENNSGKSVSLNYGEIITWLTQTEIKQKVYKYSQMINISDYINRWIFVITNNR 819  
 Qy 176 MGNKIYKINGKLIDITIKVKEITGFINFSKTIITFEINKIPDTGLITSDSDINNMWIRDFYIF 235  
 Db 820 LNSEKIYINGRLIDQKPIFISNLGNIHASNNIMFKLDCRDT-----HRYTWIKYFNL 871  
 Qy 236 AKELDGKIDINILFNSLOYTNVVDYWGNDLRYNKYMYNIDYLNRYMYANS----- 287  
 Db 872 DKELNEKEIKDLVDNOSNGILKDFWGDYQYDKPYIMLNLDPNKYVDVNVNGIRGYM 931  
 Qy 288 ---RQIVFVTR-RNNDFNEGYKIIKIRGNTNDRVRRGGDILYFDMTINKKAYNLFM 342  
 Db 932 LKGRGSMVTNIYLSLRYGKFKIKKYASGNKDNIVRNDRVYINVVKNKEVRL-- 989  
 Qy 343 KNETMYADNTHSTEDIIAIGUREQTKDINDNITFQIQPMNNTYYASQIFKSNFNGENISG 402  
 Db 990 -----ATNASQAGVERKILSALETPDVGNSLQVVMKSKNDQGITNKC-KMNLQDNNGND 1042  
 Qy 403 ICSIGTYRFRGLGWDYRHNLYPTVYKQGNVYASLESTST-----THMGFVPVSE 450  
 Db 1043 IGFICGPHQF-----NNIAKLVASNWTNROIERSSRTLGCSEWEIFPDD 1085

RESULT 13

AAR95010

ID AAR95010 standard; Protein; 1296 AA.

XX AAR95010;

XX 09-JUL-1996 (first entry)

XX C. botulinum type A neurotoxin.

XX Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen.

XX Clostridium botulinum.

XX WO9612802-A1.

XX 02-MAY-1996.

XX 23-OCT-1995; 95WO-US13737.

XX 07-JUN-1995; 95US-0480604.

XX 24-OCT-1994; 94US-0329154.

XX 16-MAR-1995; 95US-0405496.

XX 14-APR-1995; 95US-0422711.

XX (OPHI-) OPHIDIAN PHARM INC.

XX Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;

XX Williams JA;

XX WPI; 1996-230603/23.

XX N-PSDB; AAT29244.

XX Fusion proteins comprising non-toxin protein and part of toxin -

XX useful to form anti-toxins against Clostridium botulinum type A, and

XX C. difficile type toxins, and to treat C. difficile intoxication,

XX partic. diarrhoea

XX Claim 4; Page 344-350; 434pp; English.

XX Clostridium botulinum type A neurotoxin (AAR95010) is processed to form





AA93311 standard; protein; 1095 AA.  
AA93311;  
04-SEP-2000 (first entry)  
A manganese superoxide dismutase (Mn-SOD) construct.  
Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
neuronal cell targeting component; NCTC; neuronal disease;  
oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
Huntington's disease; motor neurone disease;  
botulinum neurotoxin serotype B.  
Synthetic.  
Homo sapiens.  
Bacillus stearothermophilus.  
Clostridium botulinum.  
WO200028041-A1.  
18-MAY-2000.  
05-NOV-1999; 99WO-GB03699.  
05-NOV-1998; 98GB-0024282.  
(MICR-) MICROBIOLOGICAL RES AUTHORITY.  
Shone CC, Sutton JM, Hallis B, Silman N;  
WPI; 2000-376553/32.  
Novel composition, comprising superoxide dismutase linked by a  
cleavable linker to a neuronal cell targeting component useful for  
delivering superoxide dismutase to neuronal cells to treat ischemia -  
Disclosure; Page 54-57; 65pp; English.  
The present sequence represents a construct of the invention, comprising  
a mitochondrial leader sequence from human manganese superoxide  
dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker  
that can be cleaved by thrombin, and a heavy chain derived from  
botulinum neurotoxin serotype B. The specification describes a  
composition for delivery of SOD to neuronal cells. The composition  
comprises SOD linked, by a cleavable linker, to a neuronal cell  
targeting component (NCTC). This component has a domain that binds  
to a neuronal cell and a domain that translocates the SOD of the  
composition into the neuronal cell. After translocation, the linker  
is cleaved to release the SOD. The composition is useful for treating  
neuronal diseases caused or augmented by oxidative stress, such as  
ischemic stroke, trauma, Parkinson's disease, Huntington's disease and  
motor neurone diseases.

Query Match 24.3%; Score 581; DB 21; Length 1095;  
Best Local Similarity 32.9%; Pred. NO. 2.8e-33;  
Matches 161; Conservative 82; Mismatches 168; Indels 78; Gaps 21;  
3 IPNIFSYTNNLLKDLINFPYNNINDSKLSLQNKNTLVDTSQYNAEVSDEGVDQLNP 62  
640 MPFDLSIYNTDILIEFNKYNSEILNNILNLRYKDNLDLSGYGAKVEYDGVCLND 699  
63 IFPPDFKLGSSGDRGKVIYQNTQENIYVMSYSPFSFWIRNKW---VSNL--PGYT 116  
700 --KNQFKLTSSA--NSKIRVTQNTQNIIFNSVFLDFSVFWIRPKYKNDGIONYHNEYT 755  
117 IIDSVKNNSGWSTICISNLFVFLKQNESEQINSFSDISNNAPCY-NKWFVFTVNNM 175  
756 IINCMKNNSGWKISIRGNRIIWLFDINGKTSVFVEFREDISEYINRWFFVITRN- 814  
176 MGNMKIYINGKLIDITIKVELTGTFNFSKTTTFEINKIPDTGLTSDSDNIN-MWIRDFYI 234

DB 815 LNNAKIYINGKLESTNDIKDIREVIANGEIIFKLD-----GDIDRTQFIWMKFSI 865  
QY 235 FAKELDGRKDINILENSLQYTNVVKDYWGNDLRYNKEYYVNVNIDYLARIYVANS-----RQI 290  
DB 866 ENTELSQSNIEERYKIQSYSEYKDFWGNPLMYNKEYYMFNAGKNKSYIKLKKDSPVGEI 925  
QY 291 VFNTRNNN-----DFNEGYKIIIRKIGNT---NDTRVRGGDILYFD-MTINKK--- 336  
DB 926 LTRSKYNQNSKYINRYDLYIGKEFIIRK-KSNSQSINDDIVRKEDYIYLDFFNLNGEWRV 984  
QY 337 -AYNLFMKNETWYADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYIYASQ-IFKSN 394  
DB 985 YTYKFKKEE-----EKLFLAPISDSDEFYN---TIQREYDEQPTYSCQLLFKK- 1031  
QY 395 FNGENISGICISIGTYRFLGG-----DWRHNLVPTYKQGNYSALLESTST 441  
DB 1032 -DEESTDEIGLIGIHRFYESGIVFEYKDYFCISKWY-----LKEVKRKYPNKLK---GC 1082  
QY 442 HWGFVVPVSE 450  
DB 1083 NWQFIPKDE 1091

Search completed: November 7, 2002, 14:43:09  
Job time : 67 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:43:01 ; Search time 26 Seconds  
(without alignments)  
509.243 Million cell updates/sec

Title: US-09-910-186A-10  
Perfect score: 2391  
Sequence: 1 WTIPFNIFSTNNLLKDI.....NYASLLESTSHWGFPVSE 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/ias/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/ias/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/ias/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/ias/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/ias/PCUTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/ias/Backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	586.5	24.5	1296	1	US-08-480-604A-28	Sequence 28, Appl
2	586.5	24.5	1296	2	US-08-405-496A-28	Sequence 28, Appl
3	586.5	24.5	1296	4	US-08-915-136-28	Sequence 28, Appl
4	557.5	23.3	438	1	US-08-480-604A-23	Sequence 23, Appl
5	557.5	23.3	438	2	US-08-405-496A-23	Sequence 23, Appl
6	557.5	23.3	438	4	US-08-915-136-23	Sequence 23, Appl
7	557.5	23.3	462	1	US-08-480-604A-26	Sequence 26, Appl
8	557.5	23.3	462	2	US-08-405-496A-26	Sequence 26, Appl
9	557.5	23.3	462	4	US-08-915-136-26	Sequence 26, Appl
10	539.5	22.6	1169	4	US-09-255-829-20	Sequence 20, Appl
11	425	17.8	853	4	US-08-913-880C-17	Sequence 17, Appl
12	425	17.8	858	4	US-08-913-880C-16	Sequence 16, Appl
13	425	17.8	860	4	US-08-913-880C-15	Sequence 15, Appl
14	425	17.8	862	4	US-08-913-880C-14	Sequence 14, Appl
15	425	17.8	865	4	US-08-913-880C-13	Sequence 13, Appl
16	425	17.8	866	4	US-08-913-880C-12	Sequence 12, Appl
17	425	17.8	874	4	US-08-913-880C-11	Sequence 11, Appl
18	425	17.8	875	4	US-08-913-880C-10	Sequence 10, Appl
19	425	17.8	1315	4	US-08-913-880C-1	Sequence 1, Appl
20	423.5	17.7	618	1	US-08-668-381A-5	Sequence 5, Appl
21	420	17.6	452	1	US-07-618-312A-2	Sequence 2, Appl
22	420	17.6	452	1	US-07-618-312A-4	Sequence 4, Appl
23	420	17.6	452	1	US-08-280-228-2	Sequence 2, Appl
24	420	17.6	452	1	US-08-280-228-4	Sequence 4, Appl
25	397	16.6	452	1	US-08-110-786A-8	Sequence 8, Appl
26	148.5	6.2	977	4	US-09-206-942-53	Sequence 53, Appl
27	148.5	6.2	983	4	US-09-206-942-51	Sequence 51, Appl

## ALIGNMENTS

### RESULT 1

US-08-480-604A-28  
; Sequence 28, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027

Sequence 2, Appli  
Sequence 3, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 2, Appli  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 10, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 6, Appli

CLASSIFICATION: 424  
PRIOR APPLICATION DATA: US 08/329,154  
APPLICATION NUMBER: 25-OCT-1994  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA: US 08/151,907  
APPLICATION NUMBER: 02-DEC-1993  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA: US 07/985,321  
APPLICATION NUMBER: 04-DEC-1992  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA: US 07/429,791  
APPLICATION NUMBER: 31-OCT-1989  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01308  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1296 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-405-496A-28

```

FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1296 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-28

Query Match          24.5%; Score 586.5; DB 2; Length 1296;
Best Local Similarity 31.6%; Pred. 2.6e-39;
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps

QY      3 IPFNIFYTNSLAKDIINEFYNNINDSKILSLQNKRKTLVDTSGYNAEVSDEGVQLNP 62
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB      849 IPFQLSKYVDNQRLSFTFEYIKNIINTSLNRYESNLHDLRSYAGINIGSKVNPD 908

QY      63 IPPDP---FKLGSSGDRGRVIVTONENIVYSWYSFSFIWIRINKWSNLP---CYT 116
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB      909 IDKNQIQIFNELSS-----KIEVLKNAIIVSNWFSTFWIRIPKYFNSISLNNEIT 963

QY      117 ITDSVKNNSGSIGIIENFLVFLTKQNGSEQINESFYDISNNAFCY-NKWFFVTVTNM 175
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB      964 IINCENNKGWKSVLYNGELIWLTDQEIQRKVRFKYSQMINSIDYINRWLFVTITNR 1023

QY      176 NGNKIYYNGKLDTIKVKBELTGINFSKTTPEINKIPTGLTSDSNINMWIRDFVF 235
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB      1024 LNNKSIYYNGLRIDQKYISLGNTHASNNIMFKLDGCRT-----HRYIWIKYNFLF 1075

QY      236 ARELGDGINILFNSLOVTVNVKDYGNDLRNYKXYMYNIDYLNRYMYANS----- 287
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB      1076 DRELKEKEIKLDYDNQNSGIKDFGWGLDYQDKPYMLNLDPNKYVDVNVVGIRGYM 1135

QY      288 ----RQIVFNT-RNNPDFNEGKIHKTRGTNTDTRYRGGDILYFDMTINNKAYLEM 342
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB      1136 LKGPGRGSYMTNTIYLSNLSRVGTGFIIKKYASGNKDNIVRNDRVINVVVKNKEIYL-- 1193

QY      343 KNETMYADNHGETDIATGLREQPKINDINIIFIQPMNNTVYVASQIPKSFNGENISG 402
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB      1194 -----ATNASQAGEKLTSALEIPDVGNLSQVVMVMSKNDQGITKC-KMNLQDNNGND 1246

QY      403 ICISGTYFRRLGGDWYRHNLVPVTKOCNTVASILLESTS----TWHGFPDVE 450
        ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB      1247 IGFIGFHOF-----NNIAKLVASNWTNRQIERSRRTSCSEWFIPDVD 1289

```

[illegible]

RESULT 3  
US 08-915-136-28  
Sequence 28 Application US/08915136  
; Patent No. 6298960  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHVE, NISHA V.  
; APPLICANT: PADHVE, NISHA V.

APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTI-TOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/08/915,136  
FILING DATE: 02-DEC-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/480,604  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1296 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
08-915-136-28

Query Match 24.5%; Score 586.5; DB 4; Length 1296;  
Best Local Similarity 31.6%; Pred. No. 2.6e-39;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Copied from 10910186 on 05-05-2004

QY 3 IPFNFTNNSLLKADINEFNNINSKILSLQNRKNTLVDSGYNAEVSSEGDVQLNP 62  
DB 849 IPQLSKYVDNRQLSTFTTEYIKNIINTSILNRYESNHLILSRVASKINIGSKVNDP 908  
QY 63 IPFPD---FKLGSSGDEGRKGVIVTQENINIVNMYESFSISFWIRINKWVSNLP---GYT 116  
DB 909 IDKNOIQFNLESS-----KEVILKNAIVNMYENFSFWIRIRPKYNSISLNEYT 963  
QY 117 IIDSVKNSGSIIGIISNFTLKQNSDSQSINFSYDISNAPGY-NKWFVFTVNNM 175  
DB 964 IINCENNSGKWSVSLNYGELIITLQDQEIQRVVKVYQSOMINISDYINRWIFVTITNR 1023  
QY 176 MGNMKIYNGKLIDIKVYKELTGTFNSKTIITFEINKIPDTCGLTSDSDNINMWIRDFYIF 235  
DB 1024 LNNSKIYINGRLIDQKPIISNIGNHASNNIMFKLDGCRDT-----HRYIWKYFNLF 1075

QY 236 AKELDGKDINILNSLOYTNVVDYKGNDRYNKYNNIDLYNMYANS----- 287  
DB 1076 DKELNEIKDLNDYDQNSGILKDFWGLDYQDPKPYMLNLYDPNKYVDVNNVIRGYMY 1135  
QY 288 ---RQIVFNTFR-RNNDFNEGKIIKIRIRGTNTDTRVRGGDILYFDMTINNKAYNLFM 342  
DB 1136 LKPRGSGVMTNIVLNSSLYRGKFKIIVKASGKNKDNVRNDRVINNVVKNKEVRL-- 1193  
QY 343 KNETMYADNHSTEDIYAIGLEBQTKDINDNIIQIOPMNTYVYASQIFKSPNGENISG 402  
DB 1194 -----ATNASQAGVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNGND 1246  
QY 403 ICSIGTYRFRGLGGDWYRHNLYVPTVKQGNVYASLLESTST---THWGFVFPVSE 450  
DB 1247 IGFIFGTFQF-----NNIAKIVASNNWNRQIERSSRRLGCSWEIFPVD 1289

#### RESULT 4

US-08-480-604A-23  
Sequence 23, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTI-TOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,604A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 438 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-480-604A-23

Query Match 23.3%; Score 557.5; DB 1; Length 438;  
 Best Local Similarity 31.4%; Pred. No. 1.3e-37;  
 Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

16 LKDIINEYFNINDSKILSLQNRKNTLVDTSYNAEVSSEGVDQLNPFPFD---FKLGS 72  
 4 LLSFTFYIKNIINTSLNLRYESHLIDLSRYASKINIGSKVNFDPDKNQIQLFNLES 63  
 73 SGEDRGKVIYQENIVYNSMYESPSISFWIRKINWVSNLP---GYTIIDSVKNNSGHSI 129  
 64 S-----KIEVILKNAIVNSMYENFSTFWIRPKYFNSISLNNEYTIINCNNNSGKVI 118  
 130 GIISNLFVTLKQNESEISINFSYDISNAPGY-NKWPFTVTNNMGNMKIYINGKLI 188  
 119 SLNVEIITWLTQTEIKQKVYKYSQMINISDYINRWIFVTITNRLNNSKIYINGRLI 178  
 189 DTIKVKELTGINSKTIITFEINKIPDTGLTSDSDNNINWIRDFYIFAKELDGKIDNILE 248  
 179 DQKPIISNLGNIHASNIMFKLDGCRDT-----HRYIWKYFNLFDKELKEIKDLY 230  
 249 NSLQYTNVVDYNDLRYNKEYYVYVNDYLNRYMYANS-----RQIVFNTR- 295  
 231 DNQNSGILKDFWGDYLDYDKPYMLNLDPNKYVDVNNVGVYLYLKGPGSVYATTNI 290  
 296 RNNDFNEGKIIKIRKIRGTNDTRVRGGDILYFDMTINKKAYNLFMKNETWYADNHSTE 355  
 291 YLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVVKNKEYRL-----ATNASQA 342  
 356 DIVAIGLREQTQKIDNDNIIFQIOPMNTYVYASQIFKSNFNGENISGICSGTYRFLGG 415  
 343 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMLQDNNGNDIGFIGHQF----- 397

RESULT 5  
 US-08-405-496A-23  
 Sequence 23, Application US/08405496A  
 Patent No. 5919665  
 GENERAL INFORMATION:  
 APPLICANT: WILLIAMS, JAMES A.  
 TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
 TITLE OF INVENTION: NEUROTOXIN  
 NUMBER OF SEQUENCES: 30  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MEDLEN & CARROLL, LLP  
 STREET: 220 MONTGOMERY STREET, SUITE 2200  
 CITY: SAN FRANCISCO  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/405,496A  
 FILING DATE: 16-MAR-1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/329,154  
 FILING DATE: 25-OCT-1994  
 PRIOR-APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907  
 FILING DATE: 02-DEC-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/985,321  
 FILING DATE: 04-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/429,791  
 FILING DATE: 31-OCT-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: INGOLIA, DIANE E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: OPHD-01308  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 23:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 438 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-405-496A-23

Query Match 23.3%; Score 557.5; DB 2; Length 438;  
 Best Local Similarity 31.4%; Pred. No. 1.3e-37;  
 Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;  
 16 LKDIINEYFNINDSKILSLQNRKNTLVDTSYNAEVSSEGVDQLNPFPFD---FKLGS 72  
 4 LLSFTFYIKNIINTSLNLRYESHLIDLSRYASKINIGSKVNFDPDKNQIQLFNLES 63  
 73 SGEDRGKVIYQENIVYNSMYESPSISFWIRKINWVSNLP---GYTIIDSVKNNSGHSI 129  
 64 S-----KIEVILKNAIVNSMYENFSTFWIRPKYFNSISLNNEYTIINCNNNSGKVI 118  
 130 GIISNLFVTLKQNESEISINFSYDISNAPGY-NKWPFTVTNNMGNMKIYINGKLI 188  
 119 SLNVEIITWLTQTEIKQKVYKYSQMINISDYINRWIFVTITNRLNNSKIYINGRLI 178  
 189 DTIKVKELTGINSKTIITFEINKIPDTGLTSDSDNNINWIRDFYIFAKELDGKIDNILE 248  
 179 DQKPIISNLGNIHASNIMFKLDGCRDT-----HRYIWKYFNLFDKELKEIKDLY 230  
 249 NSLQYTNVVDYNDLRYNKEYYVYVNDYLNRYMYANS-----RQIVFNTR- 295  
 231 DNQNSGILKDFWGDYLDYDKPYMLNLDPNKYVDVNNVGVYLYLKGPGSVYATTNI 290  
 296 RNNDFNEGKIIKIRKIRGTNDTRVRGGDILYFDMTINKKAYNLFMKNETWYADNHSTE 355  
 291 YLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVVKNKEYRL-----ATNASQA 342  
 356 DIVAIGLREQTQKIDNDNIIFQIOPMNTYVYASQIFKSNFNGENISGICSGTYRFLGG 415  
 343 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMLQDNNGNDIGFIGHQF----- 397  
 416 DWYRHNYLVPTVKQGYASLLESTST---THWGFVPVSE 450  
 398 -----NNIAKIVASWYNNRQIERSSRTLGSWEFIPVDD 431

RESULT 6  
 US-08-915-136-23  
 Sequence 23, Application US/08915136  
 Patent No. 6290960  
 GENERAL INFORMATION:  
 APPLICANT: KINK, JOHN A.  
 APPLICANT: THALLEY, BRUCE S.  
 APPLICANT: PADHIE, NISHA V.  
 APPLICANT: FIRCA, JOSEPH R.  
 APPLICANT: STAFFORD, DOUGLAS C.  
 TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
 PREVENTION OF C. DIFFICILE DISEASE  
 NUMBER OF SEQUENCES: 32

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: MEDLEN & CARROLL, LLP  
;; STREET: 220 MONTGOMERY STREET, SUITE 2200  
;; CITY: SAN FRANCISCO  
;; STATE: CALIFORNIA  
;; COUNTRY: UNITED STATES OF AMERICA  
;; ZIP: 94104  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/915,136  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/480,604  
;; FILING DATE:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/405,496  
;; FILING DATE: 16-MAR-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/329,154  
;; FILING DATE: 25-OCT-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/161,907  
;; FILING DATE: 02-DEC-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/985,321  
;; FILING DATE: 04-DEC-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/429,791  
;; FILING DATE: 31-OCT-1989  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: INGOLIA, DIANE E.  
;; REGISTRATION NUMBER: 40,027  
;; REFERENCE/DOCKET NUMBER: OPND-01763  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 705-8410  
;; TELEFAX: (415) 397-8338  
;; INFORMATION FOR SEQ ID NO: 23:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 438 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-915-136-23

Copied from 10910186 on 05-05-2004

Query Match 23.3%; Score 557.5; DB 4; Length 438;  
Best Local Similarity 31.4%; Pred. No. 1.3e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

16 LKDIINEYFNINDSKILQNKNTLVDTSYNAEVSDEGVDQLNPIFFPD---FKLGS 72  
4 LLSTFEYIKNIINTSILNRYESNHLIDLSRYASKINIGSKVNFDPDKNQIQLFNLES 63  
73 SGDEKGVITQENIVYNSMTSESSISFWIRINKVSNLP---GYTIIDSVKNSGWSI 129  
64 S-----KIEVLKNAIVYNSMYENFTSEFWIRIPKFNISLNEYITTINGMNSGKV 118  
130 GIISNPLVFTLKONEDSEGSINFSYDISNAPGY-NKWFVFTVTNNMGNKIKYINGKLI 188  
119 SLAYGEIITWLOTQIKORVVKFSQMINISDYINRWITVTITNRLNNSKIYINGLI 178  
189 DTIKVELGTGINSKTTTEINKIPDTGLITSDSNININWIRDFYFAKELGDKDINILF 248  
179 DQKPISNLGIIHASNNIMFELDKGCRDT-----HRYTIWYFNLFDKELNEKEIKOLY 230  
249 NSLQYINWYKVDWGNLDLRNKEVYMYNIDYLRMYANS-----RQIVENTR- 295  
231 DNOSNGILKDFWGLDYLDQPKPYIMLNLDPNKYVDVNNVGIRGMYLKGPRGSVNTINI 290

QY 296 RNNDFNEGYKIIIRKIRGNTNDRVRCGDILYFDMTINNKAYNLFMKNETWYADNHSTE 355  
Db 291 YLNSSLYRGTEIKKIIKASGKNKNTVRNDRVYINVVYKNEYRL-----ATNASQA 342  
QY 356 DIVAIGLREQTKDINDNIIFQIPMNNTYYVASQIFKNGFNENISGICISGTYRFRLLGG 415  
Db 343 GVEKILSALEIFDVGNLQSVVVMKSKNDQGITNKC-KNNLQDNNNGNDIGFIFGHOF---- 397  
QY 416 DWYRHNLYVPTVKQGNYSALLESTS-----THWGFVPVSE 450  
Db 398 -----NNIAKLAVSNWYNRQIERSRSLTLCGSWEFIPVDD 431  
RESULT 7  
US-08-480-604A-26  
; Sequence 26, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPND-01763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-480-604A-26

Query Match 23.3%; Score 557.5; DB 1; Length 462;  
Best Local Similarity 31.4%; Pred. No. 1.4e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

QY 16 LKDIINEFNNINDSKILSLONRKNITVDTSYNAVESEGDVOLNPIFPD---FKLGS 72  
DB 28 LLSTFTEYIKNIINTSILNRYESNHLIDLSRYASKINIGSKVNFDPDKNQIOLFLES 87  
QY 73 SGEDRKVIVTQNEINIVNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNQWSI 129  
DB 88 S-----KIEVILKNAIVNSMYENSTFSWIRIPKYPNSISLNNEYTIINCMMENNSQWKV 142  
QY 130 GIISNPLVTLKONEDSEOSINFSYDINSNAPGY-NKFFVTVNNMMGNKIYINGKLI 188  
DB 143 SLNYGILITLQDTQIKORVVKYQSMINISDYINRWIFVTIINNRLNNSKIYINGRLI 202  
QY 189 DTIKVKELTGINFSTITTEINKIPDTGLITSDSNINNKWIRDFVIFAKELDKDINILF 248  
DB 203 DQKPISNLGNHASNNIMEKLDGCRDT-----HRYIWKYFNLFKELNEKEIKDLY 254  
QY 249 NSLOYNVKVDWGNLDRLNKYKYVNVNDILNRYMYANS-----RQIVFNTR- 295  
DB 255 DQNSNGILKDFWGDYLOQDKPYMLNLYDPNKYVDVNVGIRGMYLKGPRGSMVTINI 314  
QY 296 RNNDFNEGYKIIKIRGNTNDTRVGGDILYFDMTNNKAYNLFEMKNETMYADNHS 355  
DB 315 YLNSSLYRGTKFIKKYASGKNDIVRNDRYINVVYKNKEYRL-----ATNASQA 366  
QY 356 DIVAIGLRQTKDINDNIIFQIPMNTYVYASQIFKSNFNGENISGTCSTGTVRFRIGG 415  
DB 367 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KNNLDQNNNDIGFIFGHQF---- 421  
QY 416 DWYRHNYLPTVKQGYASLLESTST---THWGFYVSE 450  
DB 422 -----NNIAKLVASWYNRQIERSRTLCSSWEFIPVDD 455

RESULT 8

US-08-405-496A-26  
Sequence 26, Application US/08405496A  
Patent No. 5919665  
GENERAL INFORMATION:  
APPLICANT: WILLIAMS, JAMES A.  
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
TITLE OF INVENTION: NEUROTOXIN  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,496A  
FILING DATE: 16-MAR-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01308  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 462 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-403-496A-26

Query Match 23.3%; Score 557.5; DB 2; Length 462;  
Best Local Similarity 31.4%; Pred. No. 1.4e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

QY 16 LKDIINEFNNINDSKILSLONRKNITVDTSYNAVESEGDVOLNPIFPD---FKLGS 72  
DB 28 LLSTFTEYIKNIINTSILNRYESNHLIDLSRYASKINIGSKVNFDPDKNQIOLFLES 87  
QY 73 SGEDRKVIVTQNEINIVNSMYESFSISFWIRINKWVSNLP---GYTIIDSVKNNQWSI 129  
DB 88 S-----KIEVILKNAIVNSMYENSTFSWIRIPKYPNSISLNNEYTIINCMMENNSQWKV 142  
QY 130 GIISNPLVTLKONEDSEOSINFSYDINSNAPGY-NKFFVTVNNMMGNKIYINGKLI 188  
DB 143 SLNYGILITLQDTQIKORVVKYQSMINISDYINRWIFVTIINNRLNNSKIYINGRLI 202  
QY 189 DTIKVKELTGINFSTITTEINKIPDTGLITSDSNINNKWIRDFVIFAKELDKDINILF 248  
DB 203 DQKPISNLGNHASNNIMEKLDGCRDT-----HRYIWKYFNLFKELNEKEIKDLY 254  
QY 249 NSLOYNVKVDWGNLDRLNKYKYVNVNDILNRYMYANS-----RQIVFNTR- 295  
DB 255 DQNSNGILKDFWGDYLOQDKPYMLNLYDPNKYVDVNVGIRGMYLKGPRGSMVTINI 314  
QY 296 RNNDFNEGYKIIKIRGNTNDTRVGGDILYFDMTNNKAYNLFEMKNETMYADNHS 355  
DB 315 YLNSSLYRGTKFIKKYASGKNDIVRNDRYINVVYKNKEYRL-----ATNASQA 366  
QY 356 DIVAIGLRQTKDINDNIIFQIPMNTYVYASQIFKSNFNGENISGTCSTGTVRFRIGG 415  
DB 367 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KNNLDQNNNDIGFIFGHQF---- 421  
QY 416 DWYRHNYLPTVKQGYASLLESTST---THWGFYVSE 450  
DB 422 -----NNIAKLVASWYNRQIERSRTLCSSWEFIPVDD 455

RESULT 9

US-08-915-136-26  
Sequence 26, Application US/08915136  
Patent No. 6290960  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA



COUNTRY: UNITED STATES OF AMERICA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/915,136  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/480,604  
 FILING DATE:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/405,496  
 FILING DATE: 16-MAR-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/329,154  
 FILING DATE: 25-OCT-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/161,907  
 FILING DATE: 02-DEC-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/985,321  
 FILING DATE: 04-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/429,791  
 FILING DATE: 31-OCT-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: INGOLIA, DIANE E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: OPND-01763  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 26:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 462 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-915-136-26

Query Match 23.3%; Score 557.5; DB 4; Length 462;  
 Best Local Similarity 31.4%; Pred. No. 1.4e-37;  
 Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;  
 16 LKDIINEYNNINDSKILSLQNRKNTLVDTSGYNAEVSSEGDVQNPFPD---FKLGS 72  
 28 LLSTFEYIKNIINTSILNRYESNHLIDLSRYASKINIGSKVNFDPDKNQIQLFNLES 87  
 73 SGEDRGKVTITQNEINIVNSMYESFSISFWIRINKVSNLP---GYIILDSVKNNSGWSI 129  
 88 S-----KIEVLKNAIVNSMYENSTFWIRIPKIFNSISLNNEYTIINCMMNSGKV 142  
 130 GIISNPLVFTLKQNEDEQSINFSYDINNAPGY-NKFFVTVNNMGNKIYKGLI 188  
 143 SLNYGEIITVLOQTEIQRVVKFSQMINSDYINRWIFVTIINRLNLSKIYNGELI 202  
 189 DTIKVKELTGINFSTIPEINKIPDTGLITSDSDINMWRIDYIFAKELDGKDINILF 248  
 203 DQKPIISNLGNINHASNNIFKLDGCRDT-----HRYIWKYFNLFDELNEKEIKDLY 254  
 249 NSIQYNNVVKDYGNDLRYKNEYVMVNTDYLNRVYANS-----RQIVFENTR- 295  
 255 DNGSNGILKDFWGLQYDKREYVWMLNLDPNKYVDVNNVIRGVMYKLGPRGVMVTNI 314  
 296 RNNDFNEGYKIIIRIGRTNDRVGRGDILYFDMTINNKAYLNFMKNETMYADNHSTE 355  
 315 YLNSSLYRGTKPIIKYASGNKDINVRNDRVYINVVKNEYRL-----ATNASQA 366  
 356 DIYAIGLRQTRDINDNIITFQPMNNYVYASQIFKSNFNGENISGICSGTFRPLGG 415

Db 367 GVEKILSALEIPDVGNLSQVYVYMKSKNDQGITKNC-KWNLQDNNNGDIGFIFGHQF-----421  
 QY 416 DWYRHNYLVPTVKOGNTYASLESTS-----THWGFVPVSE 450  
 Db 422 -----NNIAKLVASNWNRYNRSRTIGSGWEFIPYDD 455  
 RESULT 10  
 US-09-255-829-20  
 ; Sequence 20, Application US/09255829  
 ; Patent No. 6461617  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Shone, Clifford Charles  
 ; APPLICANT: Quinn, Conrad Padraig  
 ; APPLICANT: Foster, Keith Alan  
 ; TITLE OF INVENTION: Recombinant Toxin Fragments  
 ; NUMBER OF SEQUENCES: 29  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.  
 ; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
 ; CITY: WASHINGTON  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20005-3934  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/255,829  
 ; FILING DATE: 23-FEB-1999  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/GB97/02273  
 ; FILING DATE: 22-AUG-1997  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/782,893  
 ; FILING DATE: 27-DEC-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: ESMOND, ROBERT W.  
 ; REGISTRATION NUMBER: 32,893  
 ; REFERENCE/DOCKET NUMBER: 1581.0130002  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-371-2600  
 ; TELEFAX: 202-371-2540  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1169 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-09-255-829-20

Query Match 22.6%; Score 539.5; DB 4; Length 1169;  
 Best Local Similarity 37.4%; Pred. No. 1.5e-35;  
 Matches 130; Conservative 64; Mismatches 117; Indels 37; Gaps 12;  
 QY 3 IFNIFSTNNSLKDIINEYNNINDSKILSLQNRKNTLVDTSGYNAEVSSEGDVQNP 62  
 Db 836 MPFDSTVNDTILFEMNKYNSEILNLIINRYKNNLIDLSGYGAKVYDGVDELND 895  
 QY 63 IFPFDKLGSSGEDRGKVIYTNQENIVNSMYESFSISFWIRINKW---VSNL--PGYT 116  
 Db 896 --KNQFKLTSSA--NSKIRVTQNNIIFNSVDFSVFWIRIPKYNKDGIQNTIHNEYT 951  
 QY 117 IIDSVKNSGSGWIGIISNPLVFTLKQNEDEQSINFSYDINNAPGY-NKFFVTVNNM-175  
 Db 952 IINCNNNSGKISIRGNRIITLIDINGKTSVFFETIRNEDISEINRFFVTIINN- 1010  
 QY 176 MGNMKIYINGKLIDITIKVKELTGINFSTIPEINKIPDTGLITSDSDIN- MWIRDFYI 234  
 Db 1011 LNNAKIYINGKLESNTDKIDREVIANGELIFKLD-----GDIRTOFIWKYFSI 1061

```

; Sequence 16, Application US/08913880C
; Patent No. 6372225
; GENERAL INFORMATION:
; APPLICANT: MATSUDA, Morihiko
; TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
; FILE REFERENCE: 216-380P
; CURRENT APPLICATION NUMBER: US/08/913,880C
; CURRENT FILING DATE: 1997-11-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 16
; LENGTH: 858
; TYPE: PRT
; ORGANISM: Clostridium tetani
; FEATURE:
; OTHER INFORMATION: Amino Acids 458 to 1315 of SEQ ID NO: 1
; US-08-913-880C-16

Query Match          17.88; Score 425; DB 4; Length 858;
Best Local Similarity 26.44; Pred. No. 2e-26;
Matches 134; Conservative 90; Mismatches 169; Indels 114; Gaps

QY      3 IPF---NIFSYTNNSLLKDIINEFYNNINDSKILSLQNRKNTLVDTSYGNAYEESEGDV 58
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      59 QLNPIPPDFLKGSGE-----DRKVVITONENIVNSMYESFSISFWIRNK-WVS 110
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      456 QLPV-----GNGKAHLVNNSESEVIVHKAMDEINDMFNFTVSWLRVFKVVAS 507
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      111 NLP-----GYTIIDSVKNN-----SGWSIGIISNLFVTLTKQNEOSEQSINFSYDISNNA 160
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      508 HLEQYGTNEYXSISSMKKHSLSIGSGWSVSLKGNLNIWLTKDSAGEVQRIIFR-DLPQKF 566
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      161 PGY--NKHFFVTNNMGNMKIYINGKLIDTITKVELAGINFSEKTIITEINKIPDGTGLI 218
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      567 NAYLANKWFITINDRLSSANLITVGLMSREITGLAIREDDNITIKLDR----- 619
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      219 TSDSDNINMIRDYIIFAKELGDKDINILFNSLQYTNVVKYDNGDLRYNKYYVYNI-- 276
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      620 -CNNNNQVVSIDKFRIFCKALNPKEIEKLYTSLTIFLADFENGLRYDTEYELIPVAS 678
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      277 -----DYL---NRYMYANSROIVENTNRNNDFNEGYKIIIKIRGNTN-DTRV 321
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      679 SSKVOLAKNIYDIWLTNAPSYTNGKLNIIYRRLYN-----GLKFIKKRYTPNNEIDSEV 733
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      322 RGGDILYFDMTINNKAY-----NLFMKNETMYADNISTEDYIAGLRQOTKQINDNI 373
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      734 KSGDPIKLYVSYNNNEHTVGYPKDGNAFNNLDRLIRVGYNAPGPIPLYKKMEAVK----- 787
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      374 IFQIQPMNTYYASQIFKSNFNGENTSGTGTGTYFRFLGGD-----WYRHNYL 423
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      788 -----LRLKTYSVQL--KLYDDKNAS-LGLVGTHTNGQIGNDPNRDILLASNWY-FNHL 837
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      424 VPTVKQGYASLLESTSHWGFVPSVE 450
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
QY      838 KDKI-----LGCDWYFVETDE 853
Ddb      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

RESULT 13
US-08-913-880C-15
; Sequence 15, Application US/08913880C
; Patent No. 6372225
; GENERAL INFORMATION:
; APPLICANT: MATSUDA, Morihiko
; TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
; FILE REFERENCE: 216-380P
; CURRENT APPLICATION NUMBER: US/08/913,880C
; CURRENT FILING DATE: 1997-11-24
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 15
; LENGTH: 860

```



Db 574 NAYLANKWVITITNDRLSSANLYINGVLMGSAEITGLGAIREDDNITLKLDL----- 626  
Qy 219 TSDSDNINMIRDFYIPAKELDGKDINILFNLSQYTNVVKDYWGNDLRYNKEYYMWNI-- 276  
Db 627 -CANNQYVSIDKFRIFCKALNPKETEKLYTSITFLRDFWGNPLRYDTEYLLIPVAS 685  
Qy 277 -----DYL--NRYWYANSROIVENTRRNNDFNEGYKIIKIRGNTN-DTRV 321  
Db 686 SSKDVQLKNITDYMILTNPASYNGKLNYYRLYN-----GLKFIKRYTPNNEIDSFV 740  
Qy 322 RGGDILYFDMTINNKA-----NLFMKNETMYADNHSTEDIYAIGLREOTKDINDNI 373  
Db 741 KSGDFIKLYSVYNNNEHIVGPKDGNAFNLDRIILRVGYNAPGIPLYKKMEAVK----- 794  
Qy 374 IFQIQPNNTYYIASQIFKSNFNGENISGICSIGTYRFRLGSD-----WYRHNYL 423  
795 -----LRDLKTSYVQL--KLYDDKNAS-LGLVGTNGOIGNDPNDRDILIASNWY-FNHL 844  
424 VPTVKOGNYASLLESTSTHMGFVPVSE 450  
845 KDKI-----LGCDWYFVPTDE 860

Arch completed: November 7, 2002, 14:47:30  
Time : 30 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:45:51 ; Search time 19 Seconds  
(without alignments)  
341.491 Million cell updates/sec

Title: US-09-910-186A-10  
Perfect score: 2391  
Sequence: 1 MTIPFNFTNNLSLKDI.....NYASLLESTHGWFPVSE 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 92612 seqs, 14418503 residues

Total number of hits satisfying chosen parameters: 92612

Minimum DB seq length: 0

Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	586.5	24.5	1295	10 US-09-726-949A-1	Sequence 1, Appli
2	546	22.8	431	8 US-08-981-087A-1	Sequence 1, Appli
3	544.5	22.8	425	10 US-09-288-326-9	Sequence 9, Appli
4	250	10.5	144	8 US-08-981-087A-2	Sequence 2, Appli
5	239	10.0	144	8 US-08-981-087A-3	Sequence 3, Appli
6	145.5	6.1	1974	9 US-09-895-913A-12	Sequence 12, Appli
7	142.5	6.0	476	10 US-09-774-414-3	Sequence 3, Appli
8	133	5.6	932	10 US-09-815-242-5578	Sequence 5578, Ap
9	133	5.6	932	10 US-09-815-242-12438	Sequence 12438, A
10	129	5.4	841	10 US-09-815-242-5779	Sequence 5779, Ap
11	129	5.4	841	10 US-09-815-242-12751	Sequence 12751, A
12	128	5.4	1430	10 US-09-740-274-6	Sequence 6, Appli
13	121.5	5.1	990	12 US-10-047-678A-7	Sequence 7, Appli
14	121	5.1	483	8 US-08-834-666A-20	Sequence 20, Appli
15	118	4.9	993	10 US-09-815-242-5809	Sequence 5809, Ap
16	118	4.9	1002	10 US-09-815-242-12899	Sequence 12899, A
17	118	4.9	1002	10 US-09-815-242-13158	Sequence 13158, A
18	116.5	4.9	515	10 US-09-925-300-1282	Sequence 1282, Ap
19	116.5	4.9	996	10 US-09-815-242-5251	Sequence 5251, Ap

20	116.5	4.9	1009	10 US-09-815-242-12141	Sequence 12141, A
21	111.5	4.7	621	10 US-09-856-247A-2	Sequence 2, Appli
22	111.5	4.7	789	10 US-09-995-587A-1	Sequence 1, Appli
23	110.5	4.6	586	10 US-09-861-451A-32	Sequence 32, Appli
24	110	4.6	815	10 US-09-815-242-5106	Sequence 5106, Ap
25	110	4.6	824	10 US-09-866-582-34	Sequence 34, Appli
26	110	4.6	1116	10 US-09-790-318-2	Sequence 2, Appli
27	108	4.5	1437	10 US-09-801-368-354	Sequence 354, App
28	107	4.5	626	10 US-09-765-272-220	Sequence 220, App
29	107	4.5	869	10 US-09-815-242-10623	Sequence 10623, A
30	106.5	4.5	1781	9 US-09-995-749A-2	Sequence 2, Appli
31	106	4.4	691	8 US-08-834-666A-12	Sequence 12, Appli
32	106	4.4	691	8 US-08-834-666A-22	Sequence 22, Appli
33	104.5	4.4	555	8 US-08-808-031A-31	Sequence 31, Appli
34	104.5	4.4	586	8 US-08-808-031A-28	Sequence 28, Appli
35	103.5	4.3	584	9 US-09-995-749A-12	Sequence 12, Appli
36	103	4.3	1093	10 US-09-801-368-392	Sequence 392, App
37	102.5	4.3	1579	10 US-09-801-368-368	Sequence 368, App
38	101.5	4.2	476	10 US-09-733-524-15	Sequence 15, Appli
39	101.5	4.2	792	10 US-09-995-587A-11	Sequence 11, Appli
40	101	4.2	283	10 US-09-846-808-7	Sequence 7, Appli
41	101	4.2	283	12 US-10-059-964-26	Sequence 26, Appli
42	101	4.2	660	10 US-09-881-752A-22	Sequence 22, Appli
43	101	4.2	1143	10 US-09-924-154-14	Sequence 14, Appli
44	100.5	4.2	886	9 US-10-002-309B-2	Sequence 2, Appli
45	100.5	4.2	1031	10 US-09-815-242-10932	Sequence 10932, A

## ALIGNMENTS

RESULT 1  
US-09-726-949A-1  
; Sequence 1, Application US/09726949A  
; Patent No. US20020137886A1  
; GENERAL INFORMATION:  
; APPLICANT: Allergan, Inc.  
; APPLICANT: Aoki, Rei-Roger  
; APPLICANT: Steward, Lance E.  
; TITLE OF INVENTION: NEUROTOXINS WITH ENHANCED TARGET  
; TITLE OF INVENTION: SPECIFICITY  
; FILE REFERENCE: 36121-20002.00  
; CURRENT APPLICATION NUMBER: US/09/726,949A  
; CURRENT FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1295  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-09-726-949A-1

Query Match 24.5%; Score 586.5; DB 10; Length 1295;  
Best Local Similarity 31.6%; Pred. No. 4.3e-37;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;  
Qy 3 IPENFSTNNLSLKDIINEYFNINDSKILQNRKNTLVDTSGYNAEYSEGDYQLNP 62  
Db 848 IPFQLSKYVDNQRLLSTFTTEKININTSLNRYESNHLIDLSRVASKINIGSKYNEFP 907  
Qy 63 IPFPD---FKLGSGEDRGKVIYVQENIYVNSYFSFIRINRWVSNLP---GYT 116  
Db 908 IDKQIQLFNLESS-----KIEVLKNAIVNYSYENFSTSFIRIPKYPNSLSLNEVT 962  
Qy 117 IIDSVKNSGWSIGIISNFIKQEDSEQISNFYSIDISNNAPGY-NKWFVTVTNM 175  
Db 963 IINCMMNSGKVSGLNTGELLTQDQEKQVVFYKYSOMINISYINRWIYFTNNR 1022  
Qy 176 MGNKKIYNGKLIDTIKVKELTGINFSKITFEINKIPDTGLITSDDNINMIRDFYIF 235  
Db 1023 LNNSKIYNGRLIQKPTISNLGNHASNIMFKLDGCRDT-----HRYIWKYFNLF 1074

10910186 on 05-05-2004

9 SYTNSSLKDIINEYFNMINDSKILSLQNKNTLVDTSYNAEVSEEGDVLNPIFFDF 68

QY 259 DYWGNDLRNKEYYMVNIDYLNRYMANS-----RQIVENTR-RNNDFNEGY 305

QY 259 DYWGNDLRNKEYYMVNIDYLNRYMANS-----RQIVENTR-RNNDFNEGY 305

Db 228 DFWGDLQYDKPYMLNLDPNKYVDVNNVGRYMYLKPRGSVMTNLYLSSLYRGT 287  
Qy 306 KILIKIRGTNDTRVGGDILYEDTANNKAYLFPKNETYADNHSFEDIYAIGLREQ 365  
Db 288 KFIKKYASGNKNIVNDRVINVVKNKEYL-----ATNASAGVEKILSALE 339  
Qy 366 TKDINDNIIQIOPMNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRHNLYVP 425  
Db 340 IPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGHQF-----NNIAK 389  
Qy 426 TVKOGNYASLLEST-----THWGFVPSVE 450  
Db 330 LVASNWNRIERSRSLTGCSEWFIIPVDD 418

## RESULT 4

US-08-981-087A-2  
Sequence 2, Application US/08981087A  
Patent No. US20020081304A1  
GENERAL INFORMATION:  
APPLICANT: Elmore, Michael J.  
APPLICANT: Mauchline, Margaret L.  
APPLICANT: Minton, Nigel P.  
APPLICANT: Pasechnik, Vladimir A.  
APPLICANT: Titball, Richard W.  
TITLE OF INVENTION: TYPE F BOTULINUM TOXIN AND USE THEREOF  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHYE P.C.  
STREET: 1100 No. US20020081304A1th Glebe Rd. 8th floor  
CITY: Arlington  
STATE: VA  
COUNTRY: USA  
ZIP: 22201-4741

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/981,087A  
FILING DATE: 27-MAY-1998  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB96/01409  
FILING DATE: 12-JUN-1996

APPLICATION NUMBER: GB 9511909.5  
FILING DATE: 12-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Crawford, Arthur R.  
REGISTRATION NUMBER: 25,327  
REFERENCE/DOCKET NUMBER: 124-688  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-816-4000  
TELEFAX: 703-816-4100

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 144 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-981-087A-2

Query Match 10.5%; Score 250; DB 8; Length 144;  
Best Local Similarity 37.7%; Pred. No. 1.3e-12;  
Matches 55; Conservative 30; Mismatches 55; Indels 6; Gaps 4;

Qy 9 SYTNLSLKDIINEVFNNDKILSKLQNKNLVDTSGTNAEVSEGOVQLNPFPDF 68  
Db 1 SYTNKILILYFNKLYKIKDKNSILDMRYENKFDISGYGNSINISGDYIYSTNRNOF 60

Qy 69 KLGSSGDRGKVIYQNTQNTVNSMYESFISFWIRKWS--NLPG-YTIDSVK-NN 124  
Db 61 GYSS--KPEVNIAQNNDIYNGRYQNTSFISFWIRKWS--NLPG-YTIDSVK-NN 118  
Qy 125 SGWSIGLISFLVFLTKQNEDESQSI 150  
Db 119 SGWKISLNYKIIWTLODTAGNNQKL 144

## RESULT 5

US-08-981-087A-3  
Sequence 3, Application US/08981087A  
Patent No. US20020081304A1  
GENERAL INFORMATION:  
APPLICANT: Elmore, Michael J.  
APPLICANT: Mauchline, Margaret L.  
APPLICANT: Minton, Nigel P.  
APPLICANT: Pasechnik, Vladimir A.  
APPLICANT: Titball, Richard W.  
TITLE OF INVENTION: TYPE F BOTULINUM TOXIN AND USE THEREOF  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHYE P.C.  
STREET: 1100 No. US20020081304A1th Glebe Rd. 8th floor  
CITY: Arlington  
STATE: VA  
COUNTRY: USA  
ZIP: 22201-4741

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/981,087A  
FILING DATE: 27-MAY-1998  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB96/01409  
FILING DATE: 12-JUN-1996

APPLICATION NUMBER: GB 9511909.5  
FILING DATE: 12-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Crawford, Arthur R.  
REGISTRATION NUMBER: 25,327  
REFERENCE/DOCKET NUMBER: 124-688  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-816-4000  
TELEFAX: 703-816-4100

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 144 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-981-087A-3

Query Match 10.0%; Score 239; DB 8; Length 144;  
Best Local Similarity 37.0%; Pred. No. 9.1e-12;  
Matches 54; Conservative 25; Mismatches 57; Indels 10; Gaps 2;

Qy 152 PSYDISNAPGY-NKFEFVTVNNMGNMKIYINGKLDTIKVKELTGINFSTITFEIN 210  
Db 2 FNYTQMISIDYINKWIFVTITNNRNSRIYINGNLIDEKISINLGDHVSNNILFKIV 61

Qy 211 KIPDTGITSDDNNWIKDFYIFAKELDKDKNILFNSLQYTNVVKDYWGNDLYNKE 270  
Db 62 GCNTRIV-----GIRYFKVDFOTELKTELETLYSDPEPDLKDFWGNLYNKR 112

Qy 271 YMYNIDYLNRYMYANSRQIVFNTR 296

Db 113 YVLLNLRDTSITQNSFNINQOR 138

RESULT 6  
US-09-895-913A-12  
; Sequence 12, Application US/09895913A  
; Patent No. US20020160456A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleanthous, Harold  
; APPLICANT: Al-Garawi, Amal  
; APPLICANT: Miller, Charles  
; APPLICANT: Tomb, Jean Francois  
; APPLICANT: Comen, Raymond P.  
; TITLE OF INVENTION: Identification of Polynucleotides  
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the  
; TITLE OF INVENTION: Genome  
; FILE REFERENCE: 06132/043002  
; CURRENT APPLICATION NUMBER: US/09/895,913A  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 08/881,227  
; PRIOR FILING DATE: 1997-06-24  
; NUMBER OF SEQ ID NOS: 368  
; SOFTWARE: FastSeq for Windows version 4.0  
; SEQ ID NO 12  
; LENGTH: 1974  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
; 09-895-913A-12

Query Match 6.1%; Score 145.5; DB 9; Length 1974;  
Best Local Similarity 21.2%; Pred. No. 0.003;  
Matches 106; Conservative 61; Mismatches 143; Indels 191; Gaps 24;  
19 IINEYNNINDSKILSLQNRKNTLVDTSGYNAEVEGDVQLNPFPDFKLGSSGDRG 78  
60 IHSLEFHTDADSKDINSQVRKQ-----FDPIPKTEPCVG 94  
79 KVIIVT-QNEIYVNSMYESFISFWIRINKWNSLPGYTIIDSV-----KNSGWS 128  
95 VCYIAPYKNQDLIGSSAF-AWSLNF-----GATVGTLLLSAQEKANNNGS 141  
129 IGIISNLFVF-----TLKQNEDEQSINFSYDISNNAFGYNNKFFVTV 171  
142 IFWGNKLLYLHGNFNATNIEFTNENNVGNPNAGGGATINFNADETINADGLNYTFQTV 201  
172 TNNMGNKIIY-----INGKLIDTK--VKELT--GINF-SKTIPE-----INKI 212  
202 ALGLQTSASQHSWANFNSKLSMEIKNSNFDFTWGFGFNFSGRITFENTFTSGWTWIN-- 259  
213 PDTGLITSDSNINWIRDFYIFAKELDGKDI-----NILFNSLQYT-----NYVKDY 260  
260 ---GATESGSSVYNNVANTDILFNSILGGIRYDLKANNIENNSQWIDVSKNVNQSS 316  
261 WNDLRYNKEYIYV-----NDYLN-----NYM 284  
317 LNGNYTFNNSRLSVKPNAAINIGDSQITALENASSLSFYNNSVANFNGTAFNGVSYLN 376  
285 AN-SQIVFNTRRNN-----DENEGKIIKIRGNTNDRVRGGDILY 328  
377 LNPQVSENFQNFNNVNTYGPFGKTFDFGNSARLI--NFKGNTN-----F 424  
329 FDMTINKKAYN-----FMKNEIYADNHSTEDIYAIGLREQTKDINDNIIFQIQP 379  
425 NQATNLRAKHINHFQGVSTFKONSTMTNLAESSQAFNALKVEGET---NFN----- 474  
380 MNTYIYASQIFKSNFNGENI 400  
475 LANSLL-----NFGNSV 488

RESULT 7  
US-09-774-414-3  
; Sequence 3, Application US/09774414

Patent No. US20020102231A1  
; GENERAL INFORMATION:  
; APPLICANT: The Institute of Physical and Chemical Research  
; TITLE OF INVENTION: Endonuclease  
; FILE REFERENCE: PH-651  
; CURRENT APPLICATION NUMBER: US/09/774,414  
; PRIOR FILING DATE: 2001-01-31  
; PRIOR APPLICATION NUMBER: 09/306,970  
; PRIOR FILING DATE: 1999-05-07  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
; US-09-774-414-3  
Query Match 6.0%; Score 142.5; DB 10; Length 476;  
Best Local Similarity 22.2%; Pred. No. 0.00089;  
Matches 81; Conservative 58; Mismatches 105; Indels 121; Gaps 20;  
QY 11 TNSLLKMD--IINEY-----FNNINDSKILSLQNRKNTLVDTSGYNAEVS 53  
Db 148 SNFYLMKYLKINKYKYLDMNMPNMYNNYNNYKGLNIK---TVLDLN--NNEFY 201  
QY 54 E-----EGDVQLNPFPDFKLGSSGDRGKVVIVTQENIYVNSMYESFISFWIRNK 107  
Db 202 DYLSGLEGGDYGTP-----GGTITFNHANDVLTIFINKRKNISILVEK 246  
QY 108 WSNLPGYTIIDSVKNN-----SGWSIGIISNFI--VFTLKQNE-DSEQSINFSYDISNN 159  
Db 247 W-----MDILKONPYVPVNAFISINIKTLAKEIFITNINLYKLYSDYKIN--QINNH 294  
QY 160 APGVNKWFFVTVN-----NMGNMKIY-----INGKLI-----DTIKVE----- 195  
Db 295 IPYVN---YLKINKLPKINMDIKNNYLAGTAADGSLSMYNNPKDTLLFKNNRPSY 351  
QY 196 -LTGINSFKTITEI-----NKIPTGLITSDSNINWIRDFYIFAKE 238  
Db 352 VISOVERKELIYVLOESFDSLSNVKVKGNRKLKDFKLTFTTDELKMFYFDFKFLPL 411  
QY 239 LOGKDINILNSQYTNVNDYKNDLRYNKEYIYVNDYLNK-----YMYANSRQIV 291  
Db 412 HDNFQFNYI--KFEFTFIKSYNNW---NRVFLGLVSEYINNINIKINDYDYIYNNK 465  
QY 292 ENRR 296  
Db 466 HNARK 470  
RESULT 8  
US-09-815-242-5578  
; Sequence 5578, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011a  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727



```
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5578
; LENGTH: 932
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5578

Query Match      5.6%   Score 133;   DB 10;   Length 932;
Best Local Similarity 17.6%   Pred. No. 0.011;
Matches 86;   Conservative 85;   Mismatches 159;   Indels 160;   Gaps 19;

6 NIFSTNNSLLKDI-----INEFNNINDSKILSLQNRKNTLVDTSYNA 50
250 NLYNAQGNIIAKGYDSETSTTTFTTNYVDQNTNIGSGFEQVAFKRENAATDKTAYPM 309
51 EYSEEGD-VOLNPIFPDFKLGSSGDEGRKVIQVTONENIVYNSWIESFSISFWIRNK-- 107
310 EYVLGNDKYSKNVVDY-----GNQKGQQLIS-STNYINN--EDLSRMTVYVYVQPK 358
108 -----WVSNLPGYTTIIDSVKNNSGWSIGIISNPL-VFTLKONEDSEQSINFSYDLSNN 159
359 KYTRETFTVTLTGKFNPDAKFKIYEVTNQNFVDSFTPTDTSKLTVDVTKFKITYSND 418
160 APGYKNKFEVTVTNMGMNMKIYI-----NGKLIDTIKVK-----ELTG 198
419 ----NKTATVDLLNGQSSDQYIIQVAYPDNSSTDNKGIDYILETQNGKSSWSNSYN 474
199 INFSTIIFTEINKIPDTGLITSDS-----DNINMWRDPIYFAKELDGKDN----- 245
475 VNGSSSTANGDQKYNLGDYVWEDTKDGKODANEKIGKGVYVILKDSNGKELDRITTDEN 534
246 -----ILFNSLQYTNVVDYWGNDLRYNKE 270
535 GKQYFTGLNGTYSEVFTLAGYPTTANAGTDDAVIDSDGLTTGVIKD--ADNMILDSG 592
271 YVMWIDYLNRYMYANSR-----QIVFNTRNNN-----D 300
593 FYKTPKYSILGDYVWYDSNKGQDSTEKGIDKVKVTLNKEGEVIGTKTDENGKYRFDN 652
301 FNEG-YKIIIRKIRGNT-----NDRVRGGDI-----LYFDMTINNKAYNLFMKNET 346
653 LDSGKYKVIKFEKPAGLTGTGNTTDDKRDAGGEVDVITDHDHDFLLDNGYY-----EET 708
347 MYADNHSTED 356
709 SDSDSDSDDSD 718

RESULT 9
US-09-815-242-12438
; Sequence 12438, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
```

```
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12438
; LENGTH: 932
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12438

Query Match      5.6%   Score 133;   DB 10;   Length 932;
Best Local Similarity 17.6%   Pred. No. 0.011;
Matches 86;   Conservative 85;   Mismatches 159;   Indels 160;   Gaps 19;

6 NIFSTNNSLLKDI-----INEFNNINDSKILSLQNRKNTLVDTSYNA 50
250 NLYNAQGNIIAKGYDSETSTTTFTTNYVDQNTNIGSGFEQVAFKRENAATDKTAYPM 309
51 EYSEEGD-VOLNPIFPDFKLGSSGDEGRKVIQVTONENIVYNSWIESFSISFWIRNK-- 107
310 EYVLGNDKYSKNVVDY-----GNQKGQQLIS-STNYINN--EDLSRMTVYVYVQPK 358
108 -----WVSNLPGYTTIIDSVKNNSGWSIGIISNPL-VFTLKONEDSEQSINFSYDLSNN 159
359 KYTRETFTVTLTGKFNPDAKFKIYEVTNQNFVDSFTPTDTSKLTVDVTKFKITYSND 418
160 APGYKNKFEVTVTNMGMNMKIYI-----NGKLIDTIKVK-----ELTG 198
419 ----NKTATVDLLNGQSSDQYIIQVAYPDNSSTDNKGIDYILETQNGKSSWSNSYN 474
199 INFSTIIFTEINKIPDTGLITSDS-----DNINMWRDPIYFAKELDGKDN----- 245
475 VNGSSSTANGDQKYNLGDYVWEDTKDGKODANEKIGKGVYVILKDSNGKELDRITTDEN 534
246 -----ILFNSLQYTNVVDYWGNDLRYNKE 270
535 GKQYFTGLNGTYSEVFTLAGYPTTANAGTDDAVIDSDGLTTGVIKD--ADNMILDSG 592
271 YVMWIDYLNRYMYANSR-----QIVFNTRNNN-----D 300
593 FYKTPKYSILGDYVWYDSNKGQDSTEKGIDKVKVTLNKEGEVIGTKTDENGKYRFDN 652
301 FNEG-YKIIIRKIRGNT-----NDRVRGGDI-----LYFDMTINNKAYNLFMKNET 346
653 LDSGKYKVIKFEKPAGLTGTGNTTDDKRDAGGEVDVITDHDHDFLLDNGYY-----EET 708
347 MYADNHSTED 356
709 SDSDSDSDDSD 718

RESULT 10
US-09-815-242-5779
; Sequence 5779, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
```

```

RESULT 11
US-09-815-242-12751
  // Sequence 12751, Application US/09815242
  // Patent NO. US20020061569A1
  // GENERAL INFORMATION:
  //   APPLICANT: Haselbeck, Robert
  //   APPLICANT: Ohlsen, Kari L.
  //   APPLICANT: Zyskind, Judith W.
  //   APPLICANT: Wall, Daniel
  //   APPLICANT: Trawick, John D.
  //   APPLICANT: Carr, Grant J.
  //   APPLICANT: Yamamoto, Robert T.
  //   APPLICANT: XU, H. Howard
  // TITLE OF INVENTION: Identification of Essential Genes in
  //   TITLE OF INVENTION: Prokaryotes
  // FILE REFERENCE: ELITRA.011A
  // CURRENT APPLICATION NUMBER: US/09/815,242
  // CURRENT FILING DATE: 2001-03-21
  // PRIOR APPLICATION NUMBER: 60/191,078
  // PRIOR FILING DATE: 2000-03-21
  // PRIOR APPLICATION NUMBER: 60/206,848
  // PRIOR FILING DATE: 2000-05-23
  // PRIOR APPLICATION NUMBER: 60/207,727
  // PRIOR FILING DATE: 2000-05-26
  // PRIOR APPLICATION NUMBER: 60/242,578
  // PRIOR FILING DATE: 2000-10-23
  // PRIOR APPLICATION NUMBER: 60/253,625
  // PRIOR FILING DATE: 2000-11-27
  // PRIOR APPLICATION NUMBER: 60/257,931
  // PRIOR FILING DATE: 2000-12-22
  // PRIOR APPLICATION NUMBER: 60/269,308
  // PRIOR FILING DATE: 2001-02-16
  // NUMBER OF SEQ ID NOS: 14110
  // SOFTWARE: FastSeq for Windows Version 4.0
  // SEQ ID NO 12751
  // LENGTH: 841
  // TYPE: PRT
  // ORGANISM: Staphylococcus aureus
US-09-815-242-12751

```

[illegible]

```

Query Match      5.4%; Score 129; DB 10; Length 841;
Best Local Similarity 17.9%; Pred. NO. 0.019;
Matches 88; Conservative 81; Mismatches 160; Indels 162; Gaps 21;

  QY      6  NIFSYYNSSLKDI-----INEYFNINDSKILSLQNRKNTLVDTSYGYNA 50
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  DB     259  NIYNAGQNLIAKGIYDSTNTTTTFYFNVDQITNVRGSEQVAFAKRKNATDTAYKM 318
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  QY     51  EYSEBGDVLNPIPFQFKLGSSGDRGVIVTQNNIYVNMYSFSISFWIRLNK--- 107
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  DB     319  EYTLGNDIYSEIIL-VDY-----GNKKAQPLIS-STNYINN---EDLSRNTAYYVNPQKN 368
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  QY     108  -----WYSNLPGGYTIIDSVKNNSGWIS---GLISNLFVFTLKQNEDESEQSINFSYDIS 157
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  DB     369  TTYTKGTFVINLTGTFKNFNAKFKLIEYTDQNGFVDSFTPTSKLDKYDQD---PDVIYS 425
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  QY     158  NNAPGYNKKWFFVYTVNNMGMGNMKYII-----NGKL---IDTIKVK-----EL 196
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  DB     426  ND-----NKTATVLMKGGTSSNKQYIIQVAYPDNBSSTONGKIDYTLDPDTKTKYSWSNSY 481
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  QY     197  TGINFSKIITFEINKIPDTGLITSDS-----DNINMWIRDFYIPAKELDGDQIN--- 245
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  DB     482  SNVNGSSGTANGDQKYNLGDYVWEDTWNKDKQDANEKGKGVYVILKDSNGHELOFTTWD 541
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  QY     246  -----ILFNSLQYTNVYVKDYWGNDLRYN 268
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  DB     542  ENKGKYQFTGLSNGTYSVEESTPAGYPTPTANVGTDDAVDSDGLTTGTGFKD---ADNMTLTD 599
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  QY     269  KFFYVYNIDYLNRYMYANSR-----QIVFVTRNNN----- 299
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
  DB     600  SGFYTPKYSGLDGYVYVDSNKGQDSTERKIGKGVKVTQNEKGEVIGTIGTKIDENKGYKXF 659
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

```

210

Qy 300 -DFNEG-YKIIIRIRNT-----NDRVRGGDILFDMTINN-----KAYNLFMKE 345  
 Db 660 DNLDSGKYKIFEFKPAGLTQTVTNTTDDKADGSEV---DVTITDHDFTLDNGYFED 716  
 Qy 346 TMYADNHSTED 356  
 Db 717 TSDSDSDSDSD 727

RESULT 12  
 US-09-740-274-6  
 ; Sequence 6, Application US/09740274  
 ; Patent No. US20020031826A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nichols, Scott E.  
 ; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
 ; FILE REFERENCE: 0357CRD  
 ; CURRENT APPLICATION NUMBER: US/09/740,274  
 ; CURRENT FILING DATE: 2000-12-19  
 ; PRIOR APPLICATION NUMBER: 09/210,361  
 ; PRIOR FILING DATE: 1998-12-11  
 ; PRIOR APPLICATION NUMBER: 09/007,999  
 ; PRIOR FILING DATE: 1998-01-16  
 ; PRIOR APPLICATION NUMBER: 08/478,704  
 ; PRIOR FILING DATE: 1995-06-07  
 ; PRIOR APPLICATION NUMBER: 09/009,620  
 ; PRIOR FILING DATE: 1998-01-20  
 ; PRIOR APPLICATION NUMBER: 08/485,243  
 ; PRIOR FILING DATE: 1995-06-07  
 ; PRIOR APPLICATION NUMBER: 09/008,172  
 ; PRIOR FILING DATE: 1998-01-16  
 ; PRIOR APPLICATION NUMBER: 08/482,711  
 ; PRIOR FILING DATE: 1995-06-07  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; EQ ID NO 6  
 ; LENGTH: 1430  
 ; TYPE: PRP  
 ; ORGANISM: streptococcus mutans  
 US-09-740-274-6

Query Match 5.4%; Score 128; DB 10; Length 1430;  
 Best Local Similarity 20.8%; Pred. No. 0.044;  
 Matches 117; Conservative 73; Mismatches 188; Indels 190; Gaps 30;  
 10 YTNNSLL-KDINEYFNND--SKILSLQNRKNT---LVDTSGYNA-----EYSEEG 56  
 850 YESSALDSQLIYEGFNFQDFTYKDSYTNKKIAQNVQLFKSWGTSEMAPQYVSSD 909  
 57 DVQNLNPPDFKLGSGGDRGVIVTONENIVNMYSEFS-----ISFWIR 104  
 910 GSFLEDSI-----IQNGYAFEDRYDLAKSKNNK--YGSQODMINAVKALHKSQIYIADW-- 961  
 105 INKWSNLPGYTIIDSVK--NNSGWSIGIISNLFVLTQKNEDE--QSINFSYDINSNAPG 162  
 962 VPDQIYNLPGEVTVATRVNDYG-----EYKDSIKNTLYAANKSNGKD 1007  
 163 YNKFFVTVTNNMG-----NKKIYINGKLID-----TIKVELTGNF----- 201  
 1008 YQARYGGAFLSELAAPYSIFNTRQISNGKKIDPSEKITAWKAKYFNGTNILRGVGYVL 1067  
 202 ---SKTITFEI-----NKPIDTGLI-----TSDSDTNMMIRD-----F 232  
 1068 KDNASDKYFELKGNQTVLPQMTNKEASTGFVNDGNGMFTYSGYQAKNSFQDAGKMW 1127  
 233 YIF-----AKELDGKINILFNSLOY-----TNVVDYGN--DLRYNKEYMYVN 275  
 1128 YFDFNNGHMVYGLQALNGEVQYFSLNGVQLRESFLENADGSKNYFGLHNGRYXNGYISFD 1187  
 276 IDYLNRYMYANSRQIV--FNTRRNND--NEGKIIILKIRGNTNTRVRGGDILYFDMT 332  
 1188 NDSKWRIFDAGVMAVLKTINGNTQTFDDGGYQVKGAWITGSDGKKR-----YFDDG 1240

Qy 333 INNKAYNLFK--KNETMYADNHSTEDIYAGLRE-----QTKDINDN--- 372  
 Db 1241 SCMAVNRFRANDKNGDWYVLN--SDGIALVGVQTINGKTYTQDGKQIKGKIITDNGKL 1298  
 Qy 373 -----IIFQIQPMNNTYYVASQ-----IFKSN-----F 395  
 Db 1299 KYFLANSSELARNIFATDSQNNWYFEGSDGAVTGSQTAGKKLYFASDGKQVKGSEFTY 1358  
 Qy 396 NGE-----NISGICSGTYRFRGLGGDWY 418  
 Db 1359 NGKVHYVHADSGBELQVNRFEADKGNWY 1386

RESULT 13  
 US-10-047-676A-7  
 ; Sequence 7, Application US/10047676A  
 ; Patent No. US20020123105A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Qi, Fengxia  
 ; APPLICANT: Caulfield, Page W.  
 ; APPLICANT: Chen, Ping W.  
 ; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS  
 ; FILE REFERENCE: UAB-17403/22  
 ; CURRENT APPLICATION NUMBER: US/10/047,676A  
 ; CURRENT FILING DATE: 2002-03-21  
 ; PRIOR APPLICATION NUMBER: US 09/627,376  
 ; PRIOR FILING DATE: 2000-07-28  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 7  
 ; LENGTH: 990  
 ; TYPE: PRP  
 ; ORGANISM: Streptococcus mutans  
 US-10-047-676A-7

Query Match 5.1%; Score 121.5; DB 12; Length 990;  
 Best Local Similarity 21.9%; Pred. No. 0.087;  
 Matches 107; Conservative 86; Mismatches 170; Indels 125; Gaps 29;  
 Qy 12 NNSLL-----KDINEYFN--NIND--SKILSLQNRKNTLVDTSGY-----NAEVS- 53  
 Db 254 NNSLELLEKLEIKAMIQDYTDNRNIGEGNSILALEKMSQIVKANAYLRVDLYDHAEKL 313  
 Qy 54 ---BEGDVQ--LNPFPDFKLGSGGDRGVIVTONENIVNMYSEFSISFWIRNKW 108  
 Db 314 AQHTKSLONLKVLSSF-----SSAVNSQKEIKNTHKFIARYGYEQ--VPLQLLLNS- 366  
 Qy 109 VSNL---PGYTIID--SVKNNSGWSIGIISNLFVLTQK-----NEDSEQSINFSY 154  
 Db 367 TSGLGFPKGISQTEVSKQNNEDSKNOKIIEFLQKFEKALRDGKEILSDDDLKOLNFTD 426  
 Qy 155 DISNAPGYNKKFFVTY-----TNMGMNKKIYINGKLIDTI-----VKVELTGIN 200  
 Db 427 EQQISGELYCFYFNFKSKKLEVSLSGLVSSQMLGNTGFRHFKLPNTIVTRKNVKTKEIFEA 486  
 Qy 201 FSKTITEIKPIDTG-----LITSDSDNINMIRDFYIFAKELDGKIDNIFNSLOYTN 255  
 Db 487 IPNTIITQLNEVPYFGRGGMISLSKSHOLERN--YTKKEMSINDIYVATSEEL-- 543  
 Qy 256 VVKDYWGNDRYRNEY-----YMNVIDYLNRYMYAN--SRQIVENTRRNNNDFN----- 302  
 Db 544 ---YF-----YSKKYEKRVIFVMN---NMENYINGSKLLRFLLEVSNSDFQNTIPTFL 590  
 Qy 303 ---BQYK-----LIIKIRGNTNTRVRGGDILYFDMTINKKAYNLFK--NETMY 348  
 Db 591 GSLDSYHNVPALYIKDIIIRKPEWTNRKSEAKTDSLKNWLTNNNVPVPRVKYIDQIY 650  
 Qy 349 ADNHSTEDIYAI--GLRE-----QTKDI-----NDNIFIQ--IOPMNT-----YYASQ 389  
 Db 651 LDLSTIDLMLFQSIKKHSFIQLLDVHVSCTNDTEILELWVPFTRSDVNAHQIYHAQN 710  
 Qy 390 IFKSNFNG 397

Db 711 IYLEDGS 718

RESULT 14

US-08-834-666A-20  
 ; Sequence 20, Application US/08834666A  
 ; Patent No. US20020044949A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kleanthous, Harold  
 ; APPLICANT: Lissolo, Ling  
 ; APPLICANT: Tomb, Jean-Francois  
 ; APPLICANT: Miller, Charles  
 ; APPLICANT: Al-Garawi, Amal  
 ; TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and  
 ; TITLE OF INVENTION: Corresponding Polynucleotide Molecules  
 ; NUMBER OF SEQUENCES: 64  
 ; CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Clark & Elbing LLP  
 STREET: 176 Federal Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/834,666A  
 FILING DATE: 01-APR-1997  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Clark, Paul T.  
 REGISTRATION NUMBER: 30,175  
 REFERENCE/DOCKET NUMBER: 06132/038001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-428-0200  
 TELEFAX: 617-428-7045  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 483 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FRAGMENT TYPE: internal  
 FEATURE:  
 NAME/KEY: Signal Sequence  
 LOCATION: 1..16  
 OTHER INFORMATION:  
 US-08-834-666A-20

Query Match 5.1%; Score 121; DB 8; Length 483;  
 Best Local Similarity 20.8%; Pred. No. 0.04;  
 Matches 96; Conservative 69; Mismatches 174; Indels 122; Gaps 22;

QY 12 NNSLLKDIINEYFNINDSKILSLQNRKNTLVDT-SGYNAEYSECDYQLNFIFFDEKL 70  
 Db 38 TGTGIQVSRAY-ENLN-----NLLTRYNELKQTASNTSSTAQID----- 78  
 QY 71 GSSGÉDRGKIVITQNEINIVYNSMYESFSIFWIRINKWVSNLPGYTIIDSVKNSGWSIG 130  
 Db 79 -NLKESARLKTTPNSANQAVSSALSAVAMQVI---VSNLANSPLTSEYN---KIN 130  
 QY 131 IISNPLVFTLKQEDSQSINFSYD-----ISNNAPGVN-----KWFVTVTN 173  
 Db 131 AISOSLQNTL-ENKNNDLKTENDYDHLTQASTIINTLOSQCPGDGGNGKPGWGINASGN 189  
 QY 174 --NMWGNKVIYNGKLDITIK----- 192  
 Db 190 ACNIEGTFNVAIT-SMIDSAKKAADARRAPESPQPSAFNADFNKNLNOVSSVINDT 248

QY 193 VKELTGINFSKTITTFEINKIPDT----GLITSDSNINM---WIRDFYIFAKE----- 238  
 Db 249 ISYLKGDNLA-TIYNTLQTPDSKGFQSLVSRSSYSILNETQYSEFQTTTKFQHNPFR 307  
 QY 239 ----LDGKDINILENSLOTYNVVDYWGNDLRYNKRYMWNIDYLNRYWYANSRQIVENT 294  
 Db 308 SVGLINSQSNNGAMGVQVLGYKQFFGKKNFFGTRYAF-PDYNHAYIKSN---FFNS 362  
 QY 295 RRNNDNFEGYKIIIRIRGNNDTR-----VRGGDILYFDMTINKKAYNLPMKNETHYA 349  
 Db 363 ASNVFTYAGASOLLNFINGSGDKNRKVSFGIFGGTALAGTTLWLSQPMNLKTTNSAYSA 422  
 QY 350 --DNHSTEDIVAIGLEQKQKINDNIIQIQ-PMNNTVYYA 387  
 Db 423 KINNTNPFQFLFNGLRLQ--GIHGVGLGVKIPITNTNYS 461

RESULT 15

US-09-815-242-5809  
 ; Sequence 5809, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; PRIOR FILING DATE: 2001-03-21  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,849  
 ; PRIOR FILING DATE: 2000-03-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 5809  
 ; LENGTH: 993  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus aureus  
 US-09-815-242-5809

Query Match 4.9%; Score 118; DB 10; Length 993;  
 Best Local Similarity 21.3%; Pred. No. 0.16;  
 Matches 104; Conservative 83; Mismatches 214; Indels 88; Gaps 23;

QY 5 FNFSYTNNSLLKDIINEYFNINDSKILSLQNRKNTL--VDTSGYNAEYSECDYQLNLP 62  
 Db 23 FNLFLNTEQIKYSDL--QLNAQMKESILTTFENLYCTLQEIINFDDGDKKVRDAKESLLKY 80  
 QY 63 IF-----PFDKLGSGSEDRGKIVITQNEINIVYNSMYESFSIFWIRINKWVSNLPGY-T 116  
 Db 81 LIRMSTRTPFGI-LSGINLGHFVNEPTRLKYGNSIQKYKVD-----GEMLYIKLVSVIE 134  
 QY 117 IIDSVKNSGWSIGTISNLFVFTLKO-----NEDSEQSINFSYDINSNAPGVNKNWFFV--T 170  
 Db 135 SDEYQN-----LKVWNSKAHLINDRIVLNEQSALYLLNNKDTSFISKNSLLVFIKTT 190

Search completed: November 7, 2002, 14:53:54  
 Job time : 24 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:43:16 ; Search time 286 Seconds

(without alignments)  
1014.440 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTIPENFTSTNNLLKDI.....NYASLLESTTHGFPVSE 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_Main:\*

- 1: /cgn2\_6/ptodata/1/paa/pctus\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*
- 9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*
- 10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*
- 11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*
- 12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*
- 13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*
- 14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*
- 15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*
- 16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*
- 17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*
- 18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*
- 19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*
- 20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*
- 21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*
- 22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*
- 23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*
- 24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*
- 25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*
- 26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*
- 27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2391	100.0	450	20	US-09-611-419A-10
2	2391	100.0	450	23	US-09-910-186A-10
3	2386	99.8	1291	1	PCT-US97-15394-60
4	2386	99.8	1291	11	US-08-704-159-60
5	2386	99.8	1291	13	US-08-954-302-1
6	2386	99.8	1291	21	US-09-791-537-13524

7	2386	99.8	1291	21	US-09-791-537-95055
8	2386	99.8	1291	26	US-10-205-516-6
9	2386	99.8	1291	26	US-10-271-012-60
10	2386	99.8	1301	26	US-10-205-516-20
11	2317	96.9	462	1	PCT-US97-15394-62
12	2317	96.9	462	11	US-08-704-159-62
13	2317	96.9	462	26	US-10-271-012-62
14	1824	76.3	1285	21	US-09-791-537-59796
15	1823	76.2	1285	21	US-09-791-537-18595
16	882.5	36.9	1280	21	US-09-791-537-75966
17	882.5	36.9	1280	21	US-09-791-537-112792
18	867.5	36.3	451	20	US-09-611-419A-12
19	867.5	36.3	1275	21	US-09-791-537-33117
20	867.5	36.3	1276	1	PCT-US97-15394-66
21	867.5	36.3	1276	11	US-08-704-159-66
22	867.5	36.3	1276	19	US-09-547-188-8
23	867.5	36.3	1276	21	US-09-791-537-13527
24	867.5	36.3	1276	21	US-09-791-537-93962
25	867.5	36.3	1276	26	US-10-205-516-8
26	867.5	36.3	1276	26	US-10-271-012-66
27	867.5	36.3	1286	26	US-10-205-516-22
28	864.5	36.2	451	23	US-09-910-186A-12
29	801.5	33.5	451	1	PCT-US97-15394-68
30	801.5	33.5	451	11	US-08-704-159-68
31	801.5	33.5	451	26	US-10-271-012-68
32	604.5	25.3	1297	1	PCT-US97-15394-77
33	604.5	25.3	1297	11	US-08-704-159-77
34	604.5	25.3	1297	19	US-09-547-188-14
35	604.5	25.3	1297	21	US-09-791-537-14255
36	604.5	25.3	1297	21	US-09-791-537-152350
37	604.5	25.3	1297	26	US-10-205-516-14
38	604.5	25.3	1297	26	US-10-271-012-77
39	604.5	25.3	1307	26	US-10-205-516-28
40	602.5	25.2	1252	26	US-10-205-516-10
41	602.5	25.2	1262	26	US-10-205-516-24
42	595.5	24.9	1250	1	PCT-US97-15394-50
43	595.5	24.9	1250	11	US-08-704-159-50
44	595.5	24.9	1250	26	US-10-271-012-50
45	594.5	24.9	449	23	US-09-910-186A-14

#### ALIGNMENTS

RESULT 1  
US-09-611-419A-10  
; Sequence 10, Application US/09611419A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Leonard A.  
; APPLICANT: Byrne, Michael P.  
; APPLICANT: Middlebrook, John L.  
; APPLICANT: Lapenotiere, Hugh  
; APPLICANT: Clayton, Michael A.  
; APPLICANT: Brown, Douglas R.  
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM  
; TITLE OF INVENTION: NEUROTOXIN  
; FILE REFERENCE: A33626 067252.0105  
; CURRENT APPLICATION NUMBER: US/09/611,419A  
; CURRENT FILING DATE: 2000-07-06  
; PRIOR APPLICATION NUMBER: PCT/US00/12890  
; PRIOR FILING DATE: 2000-05-12  
; PRIOR APPLICATION NUMBER: 60/133,865  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,866  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,867  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,868  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,869  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/146,192  
; PRIOR FILING DATE: 1999-07-29

```

; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Encoded polypeptide of a synthetic construct based
; OTHER INFORMATION: on BONTC HC
US-09-611-419A-10

```

```

Query Match      100.0%; Score 2391; DB 20; Length 450;
Best Local Similarity 100.0%; Pred. No. 2.8e-212;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 MTIPFNFSYTNNSLLKDIINEYFNINNSKILSLQNRKNTLVDTSGYNAEVSSEGDVQL 60
1 MTIPFNFSYTNNSLLKDIINEYFNINNSKILSLQNRKNTLVDTSGYNAEVSSEGDVQL 60
61 NPIPPDFKLGSSGDRGKVIYVQENIVYNSMYESFSFWIRINKWYNSNLPGYTIIDS 120
61 NPIPPDFKLGSSGDRGKVIYVQENIVYNSMYESFSFWIRINKWYNSNLPGYTIIDS 120
121 VKNNSGWSIGIISNLFVTLKQNEDESEQSFYDINNSNAPGYNKWFVYVTNNMGNNK 180
121 VKNNSGWSIGIISNLFVTLKQNEDESEQSFYDINNSNAPGYNKWFVYVTNNMGNNK 180
181 IYINGKLIDITKYKELTGINFSTIIFKINIPDTGLITSDSDNINMWIRDFYIFAKELD 240
181 IYINGKLIDITKYKELTGINFSTIIFKINIPDTGLITSDSDNINMWIRDFYIFAKELD 240
241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYVMNIDYLNRYMYANSRQIVFTRNNND 300
241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYVMNIDYLNRYMYANSRQIVFTRNNND 300
301 FNEGKIIIRIGRNTNDRVRGDIIFYDWTINNKAYNLFKNETMYADNHSIEDIYAI 360
301 FNEGKIIIRIGRNTNDRVRGDIIFYDWTINNKAYNLFKNETMYADNHSIEDIYAI 360
361 GLRQTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 420
361 GLRQTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 420
421 NYLVPTVKQGYNASLLESTSTHWGFVPVSE 450
421 NYLVPTVKQGYNASLLESTSTHWGFVPVSE 450

```

```

RESULT 2
US-09-910-186A-10
Sequence 10, Application US/09910186A
GENERAL INFORMATION:
APPLICANT: U.S. Army Medical Research & Materiel Command
TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
FILE REFERENCE: A33626-A 067252.0107
CURRENT FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: US/09/910,186A
PRIOR FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: PCT/US00/12890
PRIOR FILING DATE: 2000-05-12
PRIOR APPLICATION NUMBER: 09/611,419
PRIOR FILING DATE: 2000-07-06
PRIOR APPLICATION NUMBER: 60/133,865
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 60/133,866
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 60/133,867
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 60/133,868
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 60/133,869
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 60/133,873

```

```

; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 08/123,975
; PRIOR FILING DATE: 1993-09-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
; OTHER INFORMATION: sequence
US-09-910-186A-10

```

```

Query Match      100.0%; Score 2391; DB 23; Length 450;
Best Local Similarity 100.0%; Pred. No. 2.8e-212;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 MTIPFNFSYTNNSLLKDIINEYFNINNSKILSLQNRKNTLVDTSGYNAEVSSEGDVQL 60
Db 1 MTIPFNFSYTNNSLLKDIINEYFNINNSKILSLQNRKNTLVDTSGYNAEVSSEGDVQL 60
61 NPIPPDFKLGSSGDRGKVIYVQENIVYNSMYESFSFWIRINKWYNSNLPGYTIIDS 120
61 NPIPPDFKLGSSGDRGKVIYVQENIVYNSMYESFSFWIRINKWYNSNLPGYTIIDS 120
121 VKNNSGWSIGIISNLFVTLKQNEDESEQSFYDINNSNAPGYNKWFVYVTNNMGNNK 180
121 VKNNSGWSIGIISNLFVTLKQNEDESEQSFYDINNSNAPGYNKWFVYVTNNMGNNK 180
181 IYINGKLIDITKYKELTGINFSTIIFKINIPDTGLITSDSDNINMWIRDFYIFAKELD 240
181 IYINGKLIDITKYKELTGINFSTIIFKINIPDTGLITSDSDNINMWIRDFYIFAKELD 240
241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYVMNIDYLNRYMYANSRQIVFTRNNND 300
241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYVMNIDYLNRYMYANSRQIVFTRNNND 300
301 FNEGKIIIRIGRNTNDRVRGDIIFYDWTINNKAYNLFKNETMYADNHSIEDIYAI 360
301 FNEGKIIIRIGRNTNDRVRGDIIFYDWTINNKAYNLFKNETMYADNHSIEDIYAI 360
361 GLRQTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 420
361 GLRQTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 420
421 NYLVPTVKQGYNASLLESTSTHWGFVPVSE 450
421 NYLVPTVKQGYNASLLESTSTHWGFVPVSE 450

```

```

RESULT 3
PCT-US97-15394-60
Sequence 60, Application PC/TUS9715394
GENERAL INFORMATION:
APPLICANT: Williams, James A.
APPLICANT: Thalley, Bruce S.
TITLE OF INVENTION: Multivalent Vaccine For Clostridium
TITLE OF INVENTION: Botulinum Neurotoxin
NUMBER OF SEQUENCES: 82
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medien & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: PCT/US97/15394

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-02304

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 1291 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

US-09-910-186a-10

Query Match 99.8%; Score 2386; DB 1; Length 1291;

Best Local Similarity 100.0%; Pred. No. 3.9e-211;

Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 TIFNIFSYTNNLSLKDIIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGDVQLN 61

843 TIFNIFSYTNNLSLKDIIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGDVQLN 902

62 PIFPDEFKLGSSGDEGRGVIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121

903 PIFPDEFKLGSSGDEGRGVIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962

122 KNSGWSIGIISNFIIVFTLKQNEDESEQSINFSYDISNAPGYNKWFFVTYTNMGMNMI 181

963 KNSGWSIGIISNFIIVFTLKQNEDESEQSINFSYDISNAPGYNKWFFVTYTNMGMNMI 1022

182 YINGKLIDITIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241

1023 YINGKLIDITIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082

242 KDINILFNSLOQTVNVKDYWGNDLRNKEYYMWNIIDYLNRYMYANSRQIVFNTRRNNDF 301

1083 KDINILFNSLOQTVNVKDYWGNDLRNKEYYMWNIIDYLNRYMYANSRQIVFNTRRNNDF 1142

302 NEGKILIKRIKRGNTNDRVRGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIYAIG 361

1143 NEGKILIKRIKRGNTNDRVRGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIYAIG 1202

362 LREQTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISGICSGITGRFLGSDWYRHN 421

1203 LREQTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISGICSGITGRFLGSDWYRHN 1262

422 YLVPVTKQGNYSALLESSTHGWFPVSE 450

1263 YLVPVTKQGNYSALLESSTHGWFPVSE 1291

RESULT 4

US-08-704-159-60

Sequence 60, Application US/08704159

GENERAL INFORMATION:

APPLICANT: Williams, James A.

APPLICANT: Thalley, Bruce S.

TITLE OF INVENTION: Multivalent Vaccine for Clostridium

TITLE OF INVENTION: Botulinum Neurotoxin

NUMBER OF SEQUENCES: 82

CORRESPONDENCE ADDRESS:

ADDRESS: Medlen & Carroll

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/704,159

FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-02304

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 60:

SEQUENCE CHARACTERISTICS:

LENGTH: 1291 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-704-159-60

Query Match 99.8%; Score 2386; DB 1; Length 1291;

Best Local Similarity 100.0%; Pred. No. 3.9e-211;

Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 TIFNIFSYTNNLSLKDIIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGDVQLN 61

843 TIFNIFSYTNNLSLKDIIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGDVQLN 902

62 PIFPDEFKLGSSGDEGRGVIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121

903 PIFPDEFKLGSSGDEGRGVIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962

122 KNSGWSIGIISNFIIVFTLKQNEDESEQSINFSYDISNAPGYNKWFFVTYTNMGMNMI 181

963 KNSGWSIGIISNFIIVFTLKQNEDESEQSINFSYDISNAPGYNKWFFVTYTNMGMNMI 1022

182 YINGKLIDITIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241

1023 YINGKLIDITIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082

242 KDINILFNSLOQTVNVKDYWGNDLRNKEYYMWNIIDYLNRYMYANSRQIVFNTRRNNDF 301

1083 KDINILFNSLOQTVNVKDYWGNDLRNKEYYMWNIIDYLNRYMYANSRQIVFNTRRNNDF 1142

302 NEGKILIKRIKRGNTNDRVRGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIYAIG 361

1143 NEGKILIKRIKRGNTNDRVRGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIYAIG 1202

362 LREQTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISGICSGITGRFLGSDWYRHN 421

1203 LREQTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISGICSGITGRFLGSDWYRHN 1262

422 YLVPVTKQGNYSALLESSTHGWFPVSE 450

1263 YLVPVTKQGNYSALLESSTHGWFPVSE 1291

RESULT 5

US-08-954-302-1

Sequence 1, Application US/08954302

GENERAL INFORMATION:

APPLICANT: Lance Simpson, Nikita Kiyatkin,

APPLICANT: Andrew Maksymowich

TITLE OF INVENTION: Compositions and Methods for Systemic

TITLE OF INVENTION: Delivery of Oral Vaccines and Therapeutic Agents

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESS: Jane Massey Licata, Esq.

STREET: 66 E. Main Street

CITY: Marlton

STATE: NJ



COUNTRY: USA  
ZIP: 08053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 486  
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/954,302  
FILING DATE: herewith  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane Massey Licata  
REGISTRATION NUMBER: 32,257  
REFERENCE/DOCKET NUMBER: JEFF-0164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (609) 779-2400  
TELEFAX: (609) 810-1454  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1291  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
US-08-954-302-1

Query Match 99.8%; Score 2386; DB 13; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
2 TIPNIFSYNNLSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 61  
843 TIPNIFSYNNLSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 902  
  
62 PIPFDPKLGSSGDEGRKVIIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121  
903 PIPFDPKLGSSGDEGRKVIIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962  
  
122 KNSGWSIGIISNLFVTLKQEDSEQSINFSYDINSNAPGYNKWFVTVTNMGNMKI 181  
963 KNSGWSIGIISNLFVTLKQEDSEQSINFSYDINSNAPGYNKWFVTVTNMGNMKI 1022  
  
182 YINGKLIDTIKVKELTGINFSTIPEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 241  
1023 YINGKLIDTIKVKELTGINFSTIPEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 1082  
  
242 KDINILFNSLOYTNVVKDYNGDLRYNKEYVMYVNIIDYLNRYMYANSRQIVFTRNNND 301  
1083 KDINILFNSLOYTNVVKDYNGDLRYNKEYVMYVNIIDYLNRYMYANSRQIVFTRNNND 1142  
  
302 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 361  
1143 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1202  
  
362 LREOTKDINILFOIQPMNTYVYASQIFKSNFNGENISGICSGTGYFRLLGGDWYRHN 421  
1203 LREOTKDINILFOIQPMNTYVYASQIFKSNFNGENISGICSGTGYFRLLGGDWYRHN 1262  
  
422 YLVPTVKQGNYSALLESTSTHWGFVPVSE 450  
1263 YLVPTVKQGNYSALLESTSTHWGFVPVSE 1291

RESULT 6  
US-09-791-537-13524  
; Sequence 13524, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB

; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 13524  
; LENGTH: 1291  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-09-791-537-13524

Query Match 99.8%; Score 2386; DB 21; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
2 TIPNIFSYNNLSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 61  
843 TIPNIFSYNNLSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 902  
  
62 PIPFDPKLGSSGDEGRKVIIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121  
903 PIPFDPKLGSSGDEGRKVIIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962  
  
122 KNSGWSIGIISNLFVTLKQEDSEQSINFSYDINSNAPGYNKWFVTVTNMGNMKI 181  
963 KNSGWSIGIISNLFVTLKQEDSEQSINFSYDINSNAPGYNKWFVTVTNMGNMKI 1022  
  
182 YINGKLIDTIKVKELTGINFSTIPEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 241  
1023 YINGKLIDTIKVKELTGINFSTIPEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 1082  
  
242 KDINILFNSLOYTNVVKDYNGDLRYNKEYVMYVNIIDYLNRYMYANSRQIVFTRNNND 301  
1083 KDINILFNSLOYTNVVKDYNGDLRYNKEYVMYVNIIDYLNRYMYANSRQIVFTRNNND 1142  
  
302 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 361  
1143 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1202  
  
362 LREOTKDINILFOIQPMNTYVYASQIFKSNFNGENISGICSGTGYFRLLGGDWYRHN 421  
1203 LREOTKDINILFOIQPMNTYVYASQIFKSNFNGENISGICSGTGYFRLLGGDWYRHN 1262  
  
422 YLVPTVKQGNYSALLESTSTHWGFVPVSE 450  
1263 YLVPTVKQGNYSALLESTSTHWGFVPVSE 1291

RESULT 7  
US-09-791-537-95055  
; Sequence 95055, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 95055  
; LENGTH: 1291  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum phage  
US-09-791-537-95055

Query Match 99.8%; Score 2386; DB 21; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIFPNFYSYNNLSKLDIINEYFNNINDSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 61  
 DB 843 TIFPNFYSYNNLSKLDIINEYFNNINDSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 902  
 QY 62 PIFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121  
 DB 903 PIFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962  
 QY 122 KNSGWSIGIISNFIYVTLKQNEDEQISYDINSNAPGNKWFVTVNNMGNNKI 181  
 DB 963 KNSGWSIGIISNFIYVTLKQNEDEQISYDINSNAPGNKWFVTVNNMGNNKI 1022  
 QY 182 YINGKLIDTIKVKELTGINFSTIIFKINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241  
 DB 1023 YINGKLIDTIKVKELTGINFSTIIFKINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082  
 242 KDINILFNSLQYTNVVKDYGNDLRYKKEYYVWIDYLNRYMYANSQIYVFNTRNNDF 301  
 1083 KDINILFNSLQYTNVVKDYGNDLRYKKEYYVWIDYLNRYMYANSQIYVFNTRNNDF 1142  
 302 NEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAFLFKMETYADNHSTEDIYAIG 361  
 1143 NEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAFLFKMETYADNHSTEDIYAIG 1202  
 362 LREQTKDINDNIIFOIQPMNNTYYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 421  
 1203 LREQTKDINDNIIFOIQPMNNTYYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 1262  
 422 YLVPTVKQGNYSALLESSTHGWFPVSE 450  
 1263 YLVPTVKQGNYSALLESSTHGWFPVSE 1291

RESULT 8  
 US-10-205-516-6  
 Sequence 6, Application US/10205516  
 GENERAL INFORMATION:  
 APPLICANT: Zhong, Jun  
 TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum  
 TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique  
 FILE REFERENCE: J2b1x1  
 CURRENT APPLICATION NUMBER: US/10/205,516  
 CURRENT FILING DATE: 2002-07-25  
 NUMBER OF SEQ ID NOS: 28  
 SOFTWARE: Patent in ver. 2.1  
 SEQ ID NO 6  
 LENGTH: 1291  
 TYPE: PRT  
 ORGANISM: Clostridium botulinum  
 US-10-205-516-6

Query Match 99.8%; Score 2386; DB 26; Length 1291;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
 Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIFPNFYSYNNLSKLDIINEYFNNINDSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 61  
 DB 843 TIFPNFYSYNNLSKLDIINEYFNNINDSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 902  
 QY 62 PIFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121  
 DB 903 PIFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962  
 QY 122 KNSGWSIGIISNFIYVTLKQNEDEQISYDINSNAPGNKWFVTVNNMGNNKI 181  
 DB 963 KNSGWSIGIISNFIYVTLKQNEDEQISYDINSNAPGNKWFVTVNNMGNNKI 1022  
 QY 182 YINGKLIDTIKVKELTGINFSTIIFKINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241  
 DB 1023 YINGKLIDTIKVKELTGINFSTIIFKINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082  
 242 KDINILFNSLQYTNVVKDYGNDLRYKKEYYVWIDYLNRYMYANSQIYVFNTRNNDF 301

DB 1083 KDINILFNSLQYTNVVKDYGNDLRYKKEYYVWIDYLNRYMYANSQIYVFNTRNNDF 1142  
 QY 302 NEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAFLFKMETYADNHSTEDIYAIG 361  
 DB 1143 NEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAFLFKMETYADNHSTEDIYAIG 1202  
 QY 362 LREQTKDINDNIIFOIQPMNNTYYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 421  
 DB 1203 LREQTKDINDNIIFOIQPMNNTYYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 1262  
 QY 422 YLVPTVKQGNYSALLESSTHGWFPVSE 450  
 DB 1263 YLVPTVKQGNYSALLESSTHGWFPVSE 1291

RESULT 9  
 US-10-271-012-60  
 Sequence 60, Application US/10271012  
 GENERAL INFORMATION:  
 APPLICANT: Williams, James A.  
 Thalley, Bruce S.  
 TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
 Botulinum Neurotoxin  
 NUMBER OF SEQUENCES: 82  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Medien & Carroll  
 STREET: 220 Montgomery Street, Suite 2200  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/271,012  
 FILING DATE: 15-Oct-2002  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/704,159  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ingolia, Diane E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: OPND-02304  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 60:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1291 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 60:  
 US-10-271-012-60

Query Match 99.8%; Score 2386; DB 26; Length 1291;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
 Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIFPNFYSYNNLSKLDIINEYFNNINDSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 61  
 DB 843 TIFPNFYSYNNLSKLDIINEYFNNINDSKILSLQNRKNTLVDTSQYNAEYSEEGDVLN 902  
 QY 62 PIFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121  
 DB 903 PIFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962  
 QY 122 KNSGWSIGIISNFIYVTLKQNEDEQISYDINSNAPGNKWFVTVNNMGNNKI 181

Db 963 KNSGWSIGIISFLVTLKQNEDESGSINFSYDINNAAGYKWFVTVTNMGMNMKI 1022  
 Qy 182 YINGKLIDITIKVKELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 241  
 Db 1023 YINGKLIDITIKVKELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 1082  
 Qy 242 KDINILFNSLOQYTNVVDYWGNDLRYNKEYMYMVDYLNRYMYANSRQIVFNTRRNNDF 301  
 Db 1083 KDINILFNSLOQYTNVVDYWGNDLRYNKEYMYMVDYLNRYMYANSRQIVFNTRRNNDF 1142  
 Qy 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAG 361  
 Db 1143 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAG 1202  
 Qy 362 LREQTKDINDNIIFOIPMNNTYYASQIFKSNFNGENISGICSTYRFLGGDWYRHN 421  
 Db 1203 LREQTKDINDNIIFOIPMNNTYYASQIFKSNFNGENISGICSTYRFLGGDWYRHN 1262  
 422 YLPTVKOGNYASLLESTSTHWGFPVSE 450  
 1263 YLPTVKOGNYASLLESTSTHWGFPVSE 1291  
 RESULT 10  
 Sequence 20, Application US/10205516  
 GENERAL INFORMATION:  
 APPLICANT: Zhong, Jun  
 TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum  
 FILE REFERENCE: Jb7btl  
 CURRENT APPLICATION NUMBER: US/10/205,516  
 CURRENT FILING DATE: 2002-07-25  
 NUMBER OF SEQ ID NOS: 28  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 20  
 LENGTH: 1301  
 TYPE: PRT  
 ORGANISM: Clostridium botulinum  
 us-09-910-205-516-20  
 Query Match 99.8%; Score 2386; DB 26; Length 1301;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
 Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 2 TTPNFISYTNNSLLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61  
 847 TTPNFISYTNNSLLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGVDQLN 906  
 62 PTFPDEFKLGSGEDRGKVIYQENIVYNSMYESFSISFWIRNKWWSNLPCTYTIIDSV 121  
 907 PTFPDEFKLGSGEDRGKVIYQENIVYNSMYESFSISFWIRNKWWSNLPCTYTIIDSV 966  
 122 KNSGWSIGIISNLFVTLKQNEDESGSINFSYDINNAAGYKWFVTVTNMGMNMKI 181  
 967 KNSGWSIGIISNLFVTLKQNEDESGSINFSYDINNAAGYKWFVTVTNMGMNMKI 1026  
 182 YINGKLIDITIKVKELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 241  
 1027 YINGKLIDITIKVKELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 1086  
 242 KDINILFNSLOQYTNVVDYWGNDLRYNKEYMYMVDYLNRYMYANSRQIVFNTRRNNDF 301  
 1087 KDINILFNSLOQYTNVVDYWGNDLRYNKEYMYMVDYLNRYMYANSRQIVFNTRRNNDF 1146  
 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAG 361  
 1147 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAG 1206  
 362 LREQTKDINDNIIFOIPMNNTYYASQIFKSNFNGENISGICSTYRFLGGDWYRHN 421  
 1207 LREQTKDINDNIIFOIPMNNTYYASQIFKSNFNGENISGICSTYRFLGGDWYRHN 1266

Qy 422 YLPTVKOGNYASLLESTSTHWGFPVSE 450  
 Db 1267 YLPTVKOGNYASLLESTSTHWGFPVSE 1295  
 RESULT 11  
 PCT-US97-15394-62  
 Sequence 62, Application PC/TUS9715394  
 GENERAL INFORMATION:  
 APPLICANT: Williams, James A.  
 APPLICANT: Thalley, Bruce S.  
 TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
 TITLE OF INVENTION: Botulinum Neurotoxin  
 NUMBER OF SEQUENCES: 82  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Medien & Carroll  
 STREET: 220 Montgomery Street, Suite 2200  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US97/15394  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ingolia, Diane E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: OPHD-02304  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 62:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 462 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US97-15394-62  
 Query Match 96.9%; Score 2317; DB 1; Length 462;  
 Best Local Similarity 99.8%; Pred. No. 2.1e-205;  
 Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 14 SLLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGVDQLNFIFFDFKLGSS 73  
 26 ALLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGVDQLNFIFFDFKLGSS 85  
 74 GEDRGKVIYQENIVYNSMYESFSISFWIRNKWWSNLPCTYTIIDSVKNSGWSIGIIS 133  
 86 GEDRGKVIYQENIVYNSMYESFSISFWIRNKWWSNLPCTYTIIDSVKNSGWSIGIIS 145  
 134 NLFVTLKQNEDESGSINFSYDINNAAGYKWFVTVTNMGMNMKIYINGKLIDITIKV 193  
 146 NLFVTLKQNEDESGSINFSYDINNAAGYKWFVTVTNMGMNMKIYINGKLIDITIKV 205  
 194 KELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDGKIDINILFNSLOQ 253  
 206 KELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDGKIDINILFNSLOQ 265  
 254 TNVVDYWGNDLRYNKEYMYMVDYLNRYMYANSRQIVFNTRRNNDFNEGKIIIRK 313  
 266 TNVVDYWGNDLRYNKEYMYMVDYLNRYMYANSRQIVFNTRRNNDFNEGKIIIRK 325  
 314 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAGLREQTKDINDNII 373  
 326 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAGLREQTKDINDNII 385

```

QY 374 IFQIQPNNTYYVASQIFKSNFNGENISGICSTGYFRPLGGDWYRHNLYLPTVKQNTA 433
Db 386 IFQIQPNNTYYVASQIFKSNFNGENISGICSTGYFRPLGGDWYRHNLYLPTVKQNTA 445
QY 434 SLLESTTHMGFVPVSE 450
Db 446 SLLESTTHMGFVPVSE 462

RESULT 12
US-08-704-159-62
; Sequence 62, Application US/08704159
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; TITLE OF INVENTION: Botulinum Neurotoxin
NUMBER OF SEQUENCES: 82
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/704,159
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPED-02304
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-704-159-62

Query Match 96.9%; Score 2317; DB 11; Length 462;
Best Local Similarity 99.8%; Pred. No. 2,1e-205;
Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0

14 SLKKDIINEYFNNTNDSKILSLQNRKNTLVDSGYNAEVSEGDVQLNPFPDFKLGS 73
:|||||
26 ALLKDIINEYFNNTNDSKILSLQNRKNTLVDSGYNAEVSEGDVQLNPFPDFKLGS 85
:|||||
74 GEDRKQVIVTQENIVNYSMEFSFSFWIRNKVWSNLPGYTIIDSXKNSGWSIGIIS 133
:|||||
86 GEDRKQVIVTQENIVNYSMEFSFSFWIRNKVWSNLPGYTIIDSXKNSGWSIGIIS 145
:|||||
134 NFLVFTLTKQNEDESGQSINFSDYSNNAPGYNKWFVYVTNNMGMNMKIYINGKLIDTIKV 313
:|||||
146 NFLVFTLTKQNEDESGQSINFSDYSNNAPGYNKWFVYVTNNMGMNMKIYINGKLIDTIKV 325
:|||||
194 KELTGINFSKTITTEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDNILFNLSQY 453
:|||||
206 KELTGINFSKTITTEINKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDNILFNLSQY 465
:|||||
254 TNVVKDYWGNDLRYNKZYYVNYDILNRYMYANSRQIVFENTRRNNDFNEGYKIIIKRIR 313
:|||||
266 TNVVKDYWGNDLRYNKZYYVNYDILNRYMYANSRQIVFENTRRNNDFNEGYKIIIKRIR 325
:|||||

```

QY 194 KELGINSKTIPEINIKIPDTGLITSDSDNINMWIRDFYIFAKELDKDINILFNSLOQ 253  
Db 206 KELGINSKTIPEINIKIPDTGLITSDSDNINMWIRDFYIFAKELDKDINILFNSLOQ 265  
QY 254 TNVVDYWGNDLRNKEYYVWVNDYLNRYVANSQIVFNTNRNNDNEGYKIIKRIIR 313  
Db 266 TNVVDYWGNDLRNKEYYVWVNDYLNRYVANSQIVFNTNRNNDNEGYKIIKRIIR 325  
QY 314 QNTNDRVRGGDILYFDMTINNKNAYNLFPMKNETMYADNHSTEDIYAGLRQTKDINDNI 373  
Db 326 QNTNDRVRGGDILYFDMTINNKNAYNLFPMKNETMYADNHSTEDIYAGLRQTKDINDNI 385  
QY 374 IFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGDWYRHNLYLVTVKQGNVA 433  
Db 386 IFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGDWYRHNLYLVTVKQGNVA 445  
QY 434 SILESTSTHGWFPVSE 450  
446 SILESTSTHGWFPVSE 462  
RESULT 14  
US-09-791-537-59796  
Sequence 59796, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Bionomix, Inc.  
APPLICANT: Debe, Derek  
APPLICANT: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 59796  
LENGTH: 1285  
TYPE: PRT  
ORGANISM: Clostridium botulinum  
US-09-791-537-59796  
Query Match 76.3%; Score 1824; DB 21; Length 1285;  
Best Local Similarity 77.9%; Pred. No. 5e-159;  
Matches 352; Conservative 36; Mismatches 56; Indels 8; Gaps 4;  
2 TIFPNIFSYTNNSLLKDIINEYFNNDKILSLQNKNTLVDTSGYNAEVSDEGVDQLN 61  
839 TIFPNIFSYTNNSLLKDIINEYFNNDKILSLQNKNTLVDTSGYNAEVSDEGVDQLN 898  
62 PIFPDEFKLGGSGDRGVIVTQENIVNYSMEYSISFWIRINKWVSNLPGYTIIDS 121  
899 PIFPDEFKLGGSGDRGVIVTQENIVNYSMEYSISFWIRINKWVSNLPGYTIIDS 958  
122 KNSGWSIGIISNLFVTLKQNEDESEQINFSYDISKNAAGYKWKFFVTITNNMGNMKI 181  
959 KNSGWSIGIISNLFVTLKQNEDESEQINFSYDISKNAAGYKWKFFVTITNNMGNMKI 1018  
182 YINGKLIDTIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241  
1019 YINGKLIDTIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1078  
242 KDINILFNSLOQTNVVDYWGNDLRNKEYYVWVNDYLNRYVANSQIVFNTNRNNDNF 301  
1079 KDINILFNSLOQTNVVDYWGNDLRNKEYYVWVNDYLNRYVANSQIVFNTNRNNDNF 1138  
302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKNAYNLFPMKNETMYADNHST-EDIYAI 360  
1139 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKNAYNLFPMKNETMYADNHST-EDIYAI 1193  
361 GLREQTKD-INDNIIFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGD-WY 418  
1194 GALDQPMDEIRKYGSIIFQPCNTFDYASQLFLSSNATTNRLGILSIGSYFSLGGDYWF 1253

Search completed: November 7, 2002, 14:52:41  
Job time : 294 secs

QY 419 RHNLYVPTVKQGNVASLLESTSTHGWFPVSE 450  
Db 1254 RHNLYVPTVKQGNVASLLESTSTHGWFPVSE 1285  
RESULT 15  
US-09-791-537-16595  
Sequence 16595, Application US/09791537  
GENERAL INFORMATION:  
APPLICANT: Bionomix, Inc.  
APPLICANT: Debe, Derek  
APPLICANT: Danzer, Joseph  
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
FILE REFERENCE: 261/210  
CURRENT APPLICATION NUMBER: US/09/791,537  
CURRENT FILING DATE: 2001-02-22  
NUMBER OF SEQ ID NOS: 153055  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 16595  
LENGTH: 1285  
TYPE: PRT  
ORGANISM: Clostridium botulinum phase d  
US-09-791-537-16595  
Query Match 76.2%; Score 1823; DB 21; Length 1285;  
Best Local Similarity 77.7%; Pred. No. 6.1e-159;  
Matches 351; Conservative 37; Mismatches 56; Indels 8; Gaps 4;  
2 TIFPNIFSYTNNSLLKDIINEYFNNDKILSLQNKNTLVDTSGYNAEVSDEGVDQLN 61  
839 TIFPNIFSYTNNSLLKDIINEYFNNDKILSLQNKNTLVDTSGYNAEVSDEGVDQLN 898  
62 PIFPDEFKLGGSGDRGVIVTQENIVNYSMEYSISFWIRINKWVSNLPGYTIIDS 121  
899 PIFPDEFKLGGSGDRGVIVTQENIVNYSMEYSISFWIRINKWVSNLPGYTIIDS 958  
122 KNSGWSIGIISNLFVTLKQNEDESEQINFSYDISKNAAGYKWKFFVTITNNMGNMKI 181  
959 KNSGWSIGIISNLFVTLKQNEDESEQINFSYDISKNAAGYKWKFFVTITNNMGNMKI 1018  
182 YINGKLIDTIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241  
1019 YINGKLIDTIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1078  
242 KDINILFNSLOQTNVVDYWGNDLRNKEYYVWVNDYLNRYVANSQIVFNTNRNNDNF 301  
1079 KDINILFNSLOQTNVVDYWGNDLRNKEYYVWVNDYLNRYVANSQIVFNTNRNNDNF 1138  
302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKNAYNLFPMKNETMYADNHST-EDIYAI 360  
1139 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKNAYNLFPMKNETMYADNHST-EDIYAI 1193  
361 GLREQTKD-INDNIIFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGD-WY 418  
1194 GALDQPMDEIRKYGSIIFQPCNTFDYASQLFLSSNATTNRLGILSIGSYFSLGGDYWF 1253  
QY 419 RHNLYVPTVKQGNVASLLESTSTHGWFPVSE 450  
Db 1254 RHNLYVPTVKQGNVASLLESTSTHGWFPVSE 1285

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:44:06 ; Search time 15 Seconds  
(without alignments)  
212.562 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391  
Sequence: 1 MTIPFNIFSYTNNLLKDI.....NYASLLSTSTHWGFVPVSE 450

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

28998 seqs, 7085409 residues

Total number of hits satisfying chosen parameters: 28998

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Best-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

- 1: Pending\_Patents\_AA\_New.\*
- 2: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	122	5.1	912	6	US-10-092-411A-2993
2	121.5	5.1	1335	6	US-10-092-411A-3716
3	119.5	5.0	930	6	US-10-092-411A-5314
4	117	4.9	691	6	US-10-092-411A-4675
5	116.5	4.9	1151	6	US-10-092-411A-3242
6	108	4.5	976	1	PCT-US02-05068-98
7	107.5	4.5	897	5	US-09-721-456-189
8	107.5	4.5	1007	5	US-09-721-456-187
9	107.5	4.5	1031	5	US-09-721-456-110
10	106	4.4	1010	6	US-10-092-411A-5178
11	105	4.4	10182	6	US-10-092-411A-3159
12	104.5	4.4	497	6	US-10-092-411A-5114
13	104.5	4.4	695	6	US-10-092-411A-4341
14	102	4.3	3696	6	US-10-092-411A-5080
15	100	4.2	527	6	US-10-092-411A-4465
16	100	4.2	804	6	US-10-092-411A-5218
17	99.5	4.2	670	6	US-10-092-411A-2940
18	99.5	4.2	1041	6	US-10-145-087A-498
19	99.5	4.2	1041	6	US-10-143-031A-498
20	99.5	4.2	1041	6	US-10-145-092A-498
21	99.5	4.2	1041	6	US-10-162-522A-498
22	99.5	4.2	1041	6	US-10-165-038A-498
23	99.5	4.2	1041	6	US-10-165-353-498
24	99.5	4.2	1041	6	US-10-170-481A-498
25	99.5	4.2	1041	6	US-10-172-039A-498
26	99.5	4.2	1041	6	US-10-145-016A-498

27	99.5	4.2	1041	6	US-10-145-088A-498	Sequence 498, App
28	99.5	4.2	1041	6	US-10-145-129A-498	Sequence 498, App
29	99.5	4.2	1041	6	US-10-165-353A-498	Sequence 498, App
30	99.5	4.2	4544	6	US-10-085-198-22	Sequence 22, Appl
31	99	4.1	956	6	US-10-092-411A-4452	Sequence 4452, Ap
32	98.5	4.1	422	6	US-10-092-411A-5230	Sequence 5230, Ap
33	98.5	4.1	806	6	US-10-092-411A-4314	Sequence 4314, Ap
34	97.5	4.1	428	6	US-10-092-411A-2942	Sequence 2942, Ap
35	97	4.1	509	6	US-10-092-411A-2996	Sequence 2996, Ap
36	96.5	4.0	430	6	US-10-092-411A-4712	Sequence 4712, Ap
37	96	4.0	1162	6	US-10-092-411A-4008	Sequence 4008, Ap
38	95	4.0	331	6	US-10-092-411A-3626	Sequence 3626, Ap
39	95	4.0	1151	6	US-10-060-019-31	Sequence 31, Appl
40	94.5	4.0	366	6	US-10-092-411A-4412	Sequence 4412, Ap
41	93.5	3.9	840	5	US-09-721-456-190	Sequence 190, App
42	93.5	3.9	872	5	US-09-721-456-231	Sequence 231, App
43	93.5	3.9	1183	6	US-10-092-411A-3530	Sequence 3530, Ap
44	93	3.9	267	6	US-10-092-411A-3520	Sequence 3520, Ap
45	93	3.9	478	6	US-10-271-145-2	Sequence 2, Appl

## ALIGNMENTS

### RESULT 1

US-10-092-411A-2993  
; Sequence 2993, Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: LYON DOUCETTE-STAMM et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCI  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 2993  
; LENGTH: 912  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-2993

Query Match	5.1%	Score 122;	DB 6;	Length 912;
Best Local Similarity	20.6%	Pred. No. 0.26;		
Matches	96;	Conservative	72;	Mismatches 154; Indels 144; Gaps 23;
QY	3	IPFNIFSYTN	-----SLKDIINEYFNINDSKILSLQNRKNTLVDTSGYNA--E	51
DB	374	VEFDQMPFASLKGHNPNVYTLQCI--EYKG--REVELVRKIQKNTLANKKVVYSKEE	429	
QY	52	VSEEDVOLNIFPFDFKLGSGEDRGKVIYQNTENIVY---NSMYESFSISFWIRINK	107	
DB	430	IQYGVV-----NELIDKYNNNLYKHPNREIHSFGDKYLYENE	470	
QY	108	WYSNLPFGTIIDSVKNSGWSGIGIISNLFVTL-----KQED-----SEQSINF	152	
DB	471	QSI-----MSIISKILLSNEVSGYSVEFWLNNETSLDCKEKKELLRMFSNISKISM	526	
QY	153	SYDISNNAFGYNKWFVVTNNMGNMKIYI---NGKLIDTIKVKELTGINFSTIFFEI	209	
DB	527	VY-----GAAGTGSFLINHCIFFDYDKDVIIVANTAVDNIKKK----IKLSNIKSTI	578	
QY	210	NKIPDTGLTSDSDNNINWIRDFYIFAKELGDKDINILFNSLOYTNVVKYWGNDLYNK	269	
DB	579	SK-----FLYNDKERYDLLIID---EAGTVSNKDMRILENKQFELL-----LIVGD	622	
QY	270	EYMYNIDVLYNMYVAN---SRQIVENTRRNNDFNEGYKIIKIRGINTNDTRVGGDI	326	

Db 623 NYQIESIDFGNWFIAKXVLSKNII-----NELTDMY-----RTKNDDOL 661

Qy 327 LYFDMTNNKAYNLFMKNEMWADNHS-----EDYVAGLRE- 364

Db 662 LYFKWSREKSNL---NETINMKYSTRDESIFNEFNKDEILCLNDYDGIYGINNR 718

Qy 365 --QKRDNDNIIFQIQP-----MNTYVYASQIFKSNFNGENI 400

Db 719 LQANNKNDVINGVKEYKVGDPFLFNETKNKSPILF--NNLKSII 763

RESULT 2

US-10-092-411A-3716

; Sequence 3716, Application US/10092411A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-101

; CURRENT APPLICATION NUMBER: US/10/092.411A

; PRIOR FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 09/134,001

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5676

; SEQ ID NO 3716

; LENGTH: 1335

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-10-092-411A-3716

Query Match 5.1%; Score 121.5; DB 6; Length 1335;

Best Local Similarity 19.8%; Pred. No. 0.46;

Matches 97; Conservative 74; Mismatches 193; Indels 127; Gaps 23;

10 YTNSSLKDIINEYFNINNDKILSLQNRKNTLVDSGYNAEVSSEGDVQLNPIFFDFK 69

900 YANRTEL---INKRTGNTYVLLQDGTSTL-----GWNINDVYTON-- 942

70 LGSSSEDRGKIVYTONENIVNNSM-----ESFSIFWIRINKWSNLP 114

943 IGKQTSIGKYSVPTNGLYSIANGTKNOQLAPNTLANQAFNASKAYVGK-----D 996

115 YTIIDSVKNSGWSIGIISNPLVFTLTKQNESEQSINFSYD--ISNAPGYNKWFFVYTN 173

997 LYLGTGNNRTGW---IAAKDLI---QNSTDAQSTPYNTFYVNNKSY--FYMDPTK 1046

174 NMGMNMYINGKLIDITVKELTGINF-----SKTITFEINKIPDTGLIT 219

1047 ANRYSLKPYE-QFTYVTKQKNINGVWYVQQLDGGYVWIKSTDIVKIKYVYIGMTL 1105

220 SDDSNIN-----MWIRDFYIFAKELDGKIDNILFN--SLOYTNVVKDYWN 263

1106 NNAINIQRLKRYKQVQNEPLKWSNYSQIK--NAMDTKRLANDSLKY-----Q 1154

264 DLRYNKEYVWYIDILARYM---YANSROIVENTRNNNDNEGYKIIKIR--CNTND 318

1155 FLRLDQOYL--SAQALNKLKGKGVLENQGAQSQAARKYGLNEIYLISHALVETGNGTS 1213

319 TRVRGGDIL--YFDMTNNKAYNLFMKNEMWADNHSDEIYAGLREQTKOINDNIIFQ 376

1214 QLAGGDSVSKFTTKTKGHYHVEGIGA---FDNNALVDGIRYAKNAGWTSVSKAIIIG 1270

377 IQPMNTYVYASQ--IFKSPNGENISGICSGITGYTQTTINEYTFGE-----RTASYDNTIAFSTSSG 587

1271 AKFIGNSVKAGNTLYKRWNPAN-----PGTHQYATDINWANNVNAQYKQFYDKIGE 1324

422 ----YLVPTVK 428

1325 VGKYEIETPK 1335

RESULT 3

US-10-092-411A-5314

; Sequence 5314, Application US/10092411A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-101

; CURRENT APPLICATION NUMBER: US/10/092.411A

; PRIOR FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 09/134,001

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5676

; SEQ ID NO 5314

; LENGTH: 930

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-10-092-411A-5314

Query Match 5.0%; Score 119.5; DB 6; Length 930;

Best Local Similarity 21.0%; Pred. No. 0.39;

Matches 98; Conservative 63; Mismatches 184; Indels 121; Gaps 23;

Qy 9 SYTN-----NSLLKDIINEYFNINNDKILSLQ--NRKNTLVDSGYNAEVSSEGDV 59

Db 224 SYKNIDEKISNQDELLNLPINEYFNKVRPLSTTSAPSSKRVY-----NQLAAEGSNV 278

Qy 60 LNPIFFDFKLGSGGEDRGKIVYTON--ENIVNMSYSESISFHWIRINKWSNLPGYTII 118

Db 279 NHLIKVTDQITTEYDSDGIKHAENLIYDVTE-----VD 317

Qy 119 DSVKNSGWSIGIISN-----PLVFTLTKQNESEQSINFSYDINSNAPGYNKWFFV 169

Db 318 DKVSGTMTVINDKNTVPSDLTQSFAPKIDN--SGEIIATGYDNTNKKQIYTFDYV 376

Qy 170 TVTNMNMNKKI--YI-----NGKLIDITVKELTGINFSTITFEINKIPD--TGLI 218

Db 377 DKYENIRAKHLKLSYIDKSKVPNNNTKLDVEYKLTSSVN--KTITVEYQKPNENKIANL 434

Qy 219 TSDSDNIN-----MWIRDFYIFAKELDGKIDNILFNLSLOYTNVVKDYWNLDLYNK 269

Db 435 QSMFTNIDTNHVTVEQTIYINPLRYSAKE---TNVNSGNGDEGSTLIDD---STIKVYK 489

Qy 270 EYVWNTDYLNR--YMTANSROIVENTRNNNDNEGYKIIKIRGNTNDRVVGDI-- 326

Db 490 VGDNQNLPSNRIVDISEYEDV-----TNDYVQ-----LGNNNDVNLNFGNIDS 534

Qy 327 LYFDMTI-----NNAKYNLFMKNEMWAD--NHSDEIYAGLREQTKOINDNIIFQIQ-- 378

Db 535 PYLIKVSKYDPKNDDYTTIQQVTMTQTTINEYTFGE-----RTASYDNTIAFSTSSG 587

Qy 379 -----PMNNTYVYASQIFK-----SNFNGENISGICSGITGY 409

Db 588 QGQGLDPEKTYIAGDYVWEDVDKDGQIONTNDNEKPLSNVLVLTLY 633

RESULT 4

US-10-092-411A-4675

; Sequence 4675, Application US/10092411A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-101

; CURRENT APPLICATION NUMBER: US/10/092.411A

; CURRENT FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 09/134,001

; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5676  
 ; SEQ ID NO 4675  
 ; LENGTH: 691  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-10-092-411A-4675

Query Match 4.98; Score 117; DB 6; Length 591;  
 Best Local Similarity 19.34; Pred. No. 0.4;  
 Matches 116; Conservative 84; Mismatches 176; Indels 224; Gaps 33;  
 1 MTPFNIFSYTNNSSLLKDIINEYFNINNDKIL-----LKDIINEYFNINNDKIL----- 33  
 5 MSIPNL--PTNSTMINELCLOSTINIKGEVLITELIYDDYFFKNDDEWHITAFNKFQF 62  
 34 --SLQN-----RNTL--VDTSGYNAE-----VSEE-----GDVQLNPIPPFD 67  
 63 QDSIKYRDKRNVFRIKSNLNEFLKYLFLKLVKEDWSLNFNTGAVKLNKIAKF- 121  
 68 FKLGSGEDRGVIVTQENIVYNSWYFSFISFWIRN-----KWV--SNLP-----G 114  
 122 -----FNEVYFNLSLLDCDINTLEKEWFNKLWLTENNPIKERS 159  
 115 YTIIDSVKNNSGWSIGIISNFIYFTLKQNEDESEQINSFYDISN---NAPGYNKWFFVTV 171  
 160 STIVGEDYEVKGLASFLKMYINLIKIDKREWEKDKWDIRNLEKYGLSYNK-----TL 215  
 172 TNNMGNMAYIKGLIDITKVKELTGINFESKTITFEINKIPDTGLITSDSDNINMIRD 231  
 216 TGN-----YLNFEKIESIKRELA-----KKVLKNLITGCD--TAFATAR 253  
 232 FYI-----FAKELGDKDINILFNSLQ-----YNNVVKDYWGN--DLRYNKEYYWNIDY 278  
 254 FYIRVLTRFFONSKNKTRNSNELDRCHIEAYTEFLFEVAANKHLQSTKNFVREELKT 313  
 279 LNYM-----YANSQIVF-----NTRNNNDNEGKYLIIKIRGTND-- 318  
 314 IRRFLNDITQNAIAPYQDIFLYPQPLRKHKKNSQIDYIPDFLEQLEFHNLDH 373  
 319 -----TRVRGGDILYFDMTINNKA-----YNL-----FMKNETMYADNH 352  
 374 KDLIPVWIAFKTGLRISDVL-----TLQNNCLAKYNGKYSIIITDIKTEVKGHRIPDN- 428  
 353 STEDITVAILGURE-----QTKDIN--DNIIFOI--OPMNTYTYASQIFKSNFN-----GENI-- 400  
 429 KLADIITAVLIADSKSKSTKDDNPNPNYFIAIYKGRKGMFTQHMVRAHLNHLSTKNIID 488  
 401 --SGICISIGYRFLGGDWYRHNLYPTVKQG-----NYASLLSEST 439  
 489 EQGEIHFHKTQF-----RHTYAVKLLNGGADILITQELLAHSSPEWTLRYAKLLDIT 541

RESULT 5  
 US-10-092-411A-3242  
 ; Sequence 3242, Application US/10092411A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE REFERENCE: 032796-101  
 ; CURRENT APPLICATION NUMBER: US/10/092,411A  
 ; CURRENT FILING DATE: 2002-03-07  
 ; PRIOR APPLICATION NUMBER: US 09/134,001  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5676  
 ; SEQ ID NO 3242  
 ; LENGTH: 1151  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-10-092-411A-3242  
 Query Match 4.98; Score 116.5; DB 6; Length 1151;  
 Best Local Similarity 17.98; Pred. No. 0.81;  
 Matches 97; Conservative 90; Mismatches 175; Indels 181; Gaps 27;  
 QY 3 IPFNIFSYTNNSSLLKDIINEYFNINNDKILSKILQNRKNTLV---DTSGYNAEYSEGDV 58  
 DB 72 LPSSVPPTISNTAIVSAKE-----DELIANTLQQOYTKLKTINDVKOMRQNDVDES 125  
 QY 59 QUNPIPPFD--FKLGSGEDRGV-----LVTON---ENIVYNSWYFSFIS 100  
 DB 126 ETN--FPSKNFLGTFQDTFGVDSNVATHOSSTEQPMYTSNLLFYTVDYNHVSALNFK 193  
 QY 101 FWIRINKWVSNLPGYTIIDSV-----KNNSGWSIGI----- 131  
 DB 184 FMKRINE--VGPIIFIVINQDKHNEBEITPETFKSRVEKSIDKWDIKLODTYVVSXFDH 241  
 QY 132 -----ISNFIYFTLKQNEDESEQINSFYDISNAPGYNKWFFVTVNNMGNMAYIK 184  
 DB 242 PONEIDKLSNLFVMDQHRESTDYVNRITQIFITDAQ-----YIYIONEMO----- 287  
 QY 185 GLIITIKVKELTGINFESKT--ITFEINKIPDTGLITSDSDNINMIRDYIFAKELGDKD 243  
 DB 288 -SLDITQLINE--QFEAYIQOQNO-----EVSARAQLLN-----D 322  
 QY 244 INILFNSLOYTVNVKDYWGN-----DLRYNKEYYWNIDY--LNRYMANSQIVENT 294  
 DB 323 SNQLEFNLYKQKR--KDILDNAYIMTYDMRESLRNLYLESMTATDFKVNCF-----FNK 371  
 QY 295 RNNNDNFEGYKILIKIRGTNDITRVGG-----DILYFDMTINNKAYNLFMKNE 345  
 DB 372 RKKKEEQ-----IKRLNEATQLQEVNQVROPLREDSFRLTRFNKAVNEKILNO 425  
 QY 346 -----TMADNHSTEDIYAILGUREQT--KDINDNIIFQIOPMNTYTYASQIFKSN 394  
 DB 426 EYDVVPELSIELYQTSISNTYVLTFSDEVKALNKKIENESTP-----LFEFA 475  
 QY 395 FNGENISGICS-----IGYRFLGGDWYRHNLYPTVKQGNYSALLSEST 440  
 DB 476 VNHQVYNELSDENEDRVEYDRIELNTKDSLTSHNKKHYII---HLDDSLDKLIGTE 532  
 QY 441 THW 443  
 DB 533 THF 535

RESULT 6  
 PCT-US02-05068-98  
 ; Sequence 98, Application PC/TUS0205068  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Diversa Corporation  
 ; TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY  
 ; FILE REFERENCE: 09010-108W01  
 ; CURRENT APPLICATION NUMBER: PCT/US02/05068  
 ; CURRENT FILING DATE: 2002-02-21  
 ; PRIOR APPLICATION NUMBER: US 60/270,495  
 ; PRIOR FILING DATE: 2001-02-21  
 ; PRIOR APPLICATION NUMBER: US 60/270,496  
 ; PRIOR FILING DATE: 2001-02-21  
 ; PRIOR APPLICATION NUMBER: US 60/291,122  
 ; PRIOR FILING DATE: 2001-05-14  
 ; NUMBER OF SEQ ID NOS: 321  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 98  
 ; LENGTH: 976  
 ; TYPE: PRT



ORGANISM: Bacterial  
PCT-US02-05068-98

Query Match 4.5%; Score 108; DB 1; Length 976;  
Best Local Similarity 21.8%; Pred. No. 2.4;  
Matches 101; Conservative 70; Mismatches 157; Indels 136; Gaps 27;

Qy 2 TIPNI-FSTNNSLKDI--NEVENINDSKILSLQNRKNTLVDTSGYNAEYSEGD 57  
Db 206 TVKEVDLKDASMLPEKVLHNESEIEVLPAKTKSL--TNVVDFT-----EYNEP-- 256  
Qy 58 VQLNPFP-----FDFKLGSSGDRGKIVITQNEINVINSMYSEFSIFWI-----RIN 106  
Db 257 -KLWVPQIPELFPDPTNLGSE-----INELEGLSMNDWYSDPNAEFYVYKLEEQG 310  
Qy 107 KWSNLPD-----YTIIDSVKNN-----SGWSIGIISN---FLVETLQN-----EDSE 147  
Db 311 EYLVEFGKTRKSYIENTKQEPKFGSIAINVYANGKESGLVWLKKNKLLDLESV 370  
Qy 148 QSINFSYDINNAPGYNKWFV--TVTNMMGMNKIYINGKLIDTIKVKELGFINFSKTI 205  
Db 371 DSISATINVTNKLKD-WNYTNSSVTFEV---LKKGINSNEVEII--SOLTQNSFSTEF 424  
Qy 206 T-----FEINKIPOTGLTSDSDINMWIRDFYFAKELGDKDINI-----LFNSIQYTN 255  
Db 425 TGRFVLEKIAIRVANGPESKIN-----EISRDDITITSLNPLTSSTMYTL 473  
Qy 256 VVKDYWGD-----KNETMYADNHSDEIYALGREQT-KDIN-DN 372  
Db 474 FIRSYFDTGDGVGVAEKVDYLSGLGVDVWVPLPENKSKSYHGYDVEDYDVEPDY 533  
Qy 279 LNRVYANSQIVENTFRNNDNEGYKIIKIRGNTNDTRVGRGDILYFDMTINNAY 338  
Db 534 GTLODLNMKIVL-----NEN-----GIKVVMDLVVNHSTDPHFELDAV--ENMTTNSPY 582  
Qy 339 NLFM-----KNETMYADNHSDEIYALGREQT-KDIN-DN 372  
Db 503 NYIIMSLDEPQKNHWHYKNSKGQIVWYFGLDSSMPDLNVDN 626

US-09-721-456-189  
Sequence 189, Application US/09721456  
GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

Lingner, Joachim  
Nakamura, Toru  
Chapman, Karen B.  
Morin, Gregg B.  
Harley, Calvin B.  
Andrews, William H.

TITLE OF INVENTION: Human Telomerase Catalytic Subunit

NUMBER OF SEQUENCES: 727

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/721,456  
FILING DATE: 22-Nov-2000  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,549A  
FILING DATE: 19-NOV-1997  
APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
APPLICATION NUMBER: US 08/854,050  
FILING DATE: 09-MAY-1997  
APPLICATION NUMBER: US 08/911,312  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: US 08/912,951  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: US 08/915,503  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: WO PCT/US97/17618  
FILING DATE: 01-OCT-1997  
APPLICATION NUMBER: WO PCT/US97/17885  
FILING DATE: 01-OCT-1997  
NAME: Apple, Randolph Ted  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002610US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0300  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 189:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 897 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 189:  
US-09-721-456-189

Query Match 4.5%; Score 107.5; DB 5; Length 897;  
Best Local Similarity 18.5%; Pred. No. 2.3;  
Matches 101; Conservative 80; Mismatches 174; Indels 191; Gaps 26;

Qy 7 IFSY-TNNSLLKDIINEYNNINDSKILSLQNRKNTLVDTSGY-----48  
Db 217 LFSYTDNKCVTQFINFEFFYNILPKDFLGRNKNFKQKVKYVELNKEHLTKLLEK 276  
Qy 49 --NAEVS-BEGDVQLNPPIPPDFK-----LGSSGDRGKIVITQNEINVIY-NSMYES 96  
Db 277 INTREISWQVETSAKHFFYFDHENIYVLKLLRWTFD---LVVSLRCFFYVTEQKS 333  
Qy 97 FSIWFIRINKW-----VSNLPGYTIIDSVKNNSGWSIGIISNPLVFTLKQNEDESQS 149  
Db 334 YSKTYYYRKNIDVINKMSTADLKKETLAEVQE-----KEVEENKKS 375  
Qy 150 INFSYDISNAPG-----YKNWFFVTVTNNMMGNKIVINGKLIDT-IK 192  
Db 376 LGF-----APKRLILPKTTFRPFIMFNK-----KIVNSDKTKTLTNTKLNHLM 424  
Qy 193 VKELTGINFSKTITFEINKIPD-----TGLITSNDIN---226  
Db 425 LKTLKRNMFDPGFAVENYDDVMKKYEEFVCKWKQVGPQLFFATMDIEKCYDSVNEK 484  
Qy 227 -----MWIRDFYFAKELGDKDINILFNSLOY-TNVVXDYWGNDLRYNKYYWNI 276  
Db 485 LSTFLKTKLLSSDFWIMTAQILKRNKNIIVDSKNPKKEMKDY-----FRQKQKIAL 538  
Qy 277 DYLNRYMYANSRQIVENTREN-NNDFNEGYKIIKIRGNTNDTRVGRGDILYF-----329  
Db 539 E-----GGQYPTLSVLENEQNDLNAKTLIVEAKQKN-----YFKKNDLL 579  
Qy 330 DMVINNKAYNLFKNKNTMYADNHSTEDIYALGREQTKDINDNIIFQIPMNTYVYASQ 389  
Db 580 QPVINICQYNYINFNGKFY-----KOTKGIPOGLC--VSSILUSSFYATL 622  
Qy 390 -----IFKSNFNGENISGICSTGYRFRFLGWDYRHNVLVPTVKQGNYSLLF---ST 439



SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/721,456

FILING DATE: 22-NOV-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/974,549A

FILING DATE: 19-NOV-1997

APPLICATION NUMBER: US/08/724,643

FILING DATE: 01-OCT-1996

APPLICATION NUMBER: US/08/844,419

FILING DATE: 18-APR-1997

APPLICATION NUMBER: US/08/846,017

FILING DATE: 25-APR-1997

APPLICATION NUMBER: US/08/851,843

FILING DATE: 06-MAY-1997

APPLICATION NUMBER: US/08/854,050

FILING DATE: 09-MAY-1997

APPLICATION NUMBER: US/08/911,312

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: US/08/912,951

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: US/08/915,503

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: WO PCT/US97/17618

FILING DATE: 01-OCT-1997

APPLICATION NUMBER: WO PCT/US97/17885

FILING DATE: 01-OCT-1997

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph Ted

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-0026100S

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 110:

SEQUENCE CHARACTERISTICS:

LENGTH: 1031 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 110:

09-721-456-110

Query Match 4.5%; Score 107.5; DB 5; Length 1031;  
Best Local Similarity 18.5%; Pred. No. 2.8;  
Matches 101; Conservative 80; Mismatches 174; Indels 191; Gaps 26;

7 IFSY-TNNSLLKDIINEYFNNT-----NDSKILSLONRKNLTVDTSYNAEYSEEDVQLNPI 63  
348 LFSYTDNKKVQTFNEFFYNILPKDLTGRRNKKFQKKYKVELNKKHELHKNLLEK 407  
49 --NAEVS-EEGDVQLNPIFPDFK-----LGSSGDEGRKVIIVTONENIVY-NSMYES 96  
408 INTRISWQVETSAKHYYEDHENIYVLMKLLRWIFED---LVVSLRCFFVYVTEQOKS 464  
97 FTSIWIIRNKW-----VSNLPGYTIISVKNNSGWSIGISNLFVTLKONEDSEOS 149  
465 YSKTYVYRKNIDVIMKMSIADLKKETLAEVQE-----KEVEEWKKS 506  
150 INFSDYISNAPG-----YKWFVFTVNTNMGNMKVIYINGKLIDT-IK 192  
507 LGF-----APGKLLPDKTTPRPIMTFNK-----KIVNSDRKTKLTNTKLLSHLM 555  
193 VKELTGIFNSKTIITEINKIPD-----TGLTSDSDNTIN--- 226  
556 LKTLNRNFKDPFGFAVDYDVMKKVEEYFCKWKVGQPKLFEATMDIEKCYDSVNRK 615  
227 -----MWIRDEYIFAKELDKDINLENSLOY-TNVVKDYWGNDLYNKEYYMWNI 276  
616 LSTFLTKTLSSDFWIMTAQLLKKNNIVIDSKNFKKMKDY-----FRQFKIAL 669  
277 DYLNRYMYANSQIVFNTRRN-NNDFNEGYKIIIRKIRGNTNDRVRGGDILYF----- 329

Db 670 E-----GGQYPTFLSVLENEQNDLNARKTLIVEAKORN-----YFKKDNLL 710  
QY 330 DMTINKAYNLPMKNETMYADNHSTEDIYAIGLREQTKDINDNIIFQIOPMNTYVYASQ 389  
Db 711 QPVINICQYNYINFNGKFY-----KOTKGIPOGLC--VSSILSSFYATL 753  
QY 390 -----IFKSNFNGENISICISIGTYRPLGGDWYRHHYLVPTVKQGNVYASLLE---ST 439  
Db 754 EESSLGLRDESMNPENPN-----VNLMLRLTDD-----YLLITTCENNAVLFIKLI 803  
QY 440 STHWGF 445  
Db 804 SRENGF 809

RESULT 10  
US-10-092-411A-5178  
; Sequence 5178, Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 5178  
; LENGTH: 1010  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-5178

Query Match 4.4%; Score 106; DB 6; Length 1010;  
Best Local Similarity 19.7%; Pred. No. 3.4;  
Matches 75; Conservative 60; Mismatches 100; Indels 146; Gaps 19;

QY 8 PSYNNLLKDIINEYFNNT-----NDSKILSLONRKNLTVDTSYNAEYSEEDVQLNPI 63  
Db 674 YEYENK-TKDKLNLNKKLIEKMDQKHLT-ENLTQTSKEIN--NLELKKEMQ---- 725  
QY 64 FPEDFKLGSSGDEGRKY---IVTQENI-----VYNSMYESFISFWIRINKWVSNLPGY 115  
Db 726 -----QLGFESYDQVKSADLSAQKDELEIREINLYKNKYQSYE---IEINRLKELVKCK 776  
QY 116 TIID-----SVKNNSGWSIGTISNLFVTLKONEDSEOSINFSYDISNAPGYNKWFV 169  
Db 777 KLLNLEELRQSIETN-----LKLDETNSQIATISYKIDNNSNKFNK----- 818  
QY 170 TVTNMGMNKVIYINGKLIDTKVKELTGIFNSKTIITEINKIPDTGLTSDSDNINMWI 229  
Db 819 -IKN-----LIQILDDELKVO----- 833  
QY 230 RDTYIFAKELDKDINLENSLOYTNVVKDYWGNDLYNKEYYMWIDLYNRYMYANSRO 289  
Db 834 KEIFLLSEILACK-----NDYKLTLENYVL-IYYLEKIIFQANOR 872  
QY 290 IVFNTRRNNDNFNGYKTIIRISGNTNDRVRGGDILYFDIMINNKAYNLFMKNETMYA 349  
Db 873 LSFMS-----GNRYQL-----IRRETISGLSGLEIDVED-----FHSNKRHI 911  
QY 350 DNHSTEDIY-----AIGLRE 364  
Db 912 SSLSGGEFFQASLALGLSE 932

RESULT 11

```
US-10-092-411a-3159
; Sequence 3159, Application US/10092411a
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411a
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-092-411a-3159
Query Match 4.4%; Score 105; DB 6; Length 10182;
Best Local Similarity 21.2%; Pred. No. 71;
Matches 121; Conservative 71; Mismatches 202; Indels 176; Gaps 28;
6 NPSYTNNSLLDIINEYFNNDKSLQNRKN--TLVDTSGYNAEVSSEGDVQLNP- 62
1313 NVSSVTVTRKGDLEKSSNNNNFLRWNRDYYTISSTSKIOGRNDDGLENLSPD 1372
63 ----IPFPFKLGSGGEDRGKVIYVQ----NENTVYN-----SWYSEFSISFWIRI--- 105
1373 HKSLIYKDATLGRQINTNDVLTLLQATAKNSLRNSINSNEKQLAERGSNGYSKSIIRD 1432
106 ---NKVSNLPGYTIIDSVKNNWSGIISNLFVTLKQNEDESEQSFNFSYDISSNAPG 162
1433 DGEKSVLLSNFIQVLDLVEPDNGGGRQVSHSNVT---YNEKSSIVNGOVPPANGASA 1489
163 YN--KWFVTVTN--MMG-----NMKIYINGKLIDIRY 193
1490 FNIDKVVKANAANGIMGVYKAQLAPYSPKGYIEKLQNLSTNNVINVYVPSDKV 1549
194 KE-----LTGINFSTIIFENKIPDCLITSDSDNINMTRDIYIFAKELDG 241
1550 NPSITGVNDHVTYSGETFKNTINVDNYGLNTVASTSDS-AITMTRN-----NELVG 1603
242 KGINILFNSLOITNVYK--DYWGNDLRYNKEYVMNIDYLN---RYMANSROI---VFN 293
1604 QAPNVI-NSINKIVKATDKSGNESIYS---FTVNIKPLNEKYRIITSSNQTPVRISN 1659
294 TRNNN-----DEN-----EGYKIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMK 343
1660 IQNNALSIEDQNRKSSLSMTKILGTRNVYNESNNDVRSQVYKVNKSGNNATVNV--- 1716
344 NEMVAD-----NHSDEIYA----- 359
1717 -TTTSDGTNTITVPYKHVLEVPTTTRTVRGQQFTGKGTSPNDFSLRTGCPVDAR 1775
360 -IGLEQTKDINDN-----IIF-QTOPM-NNYVYAS-----QIFKSNFNGEN 399
1776 IYVWNNQGDINSQIGRDLTLHAEIFFDGETTPRKDITVYKLSQISIPKQIYETTING-- 1833
400 ISGICSGIYRFLGWDYRHHVLYPTVKQ 429
1834 -----RFNSSGDAYPGNF-VQAVNQ 1852
RESULT 12
US-10-092-411a-5114
; Sequence 5114, Application US/10092411a
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
```

```
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411a
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 5114
; LENGTH: 497
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-092-411a-5114
Query Match 4.4%; Score 104.5; DB 6; Length 497;
Best Local Similarity 21.3%; Pred. No. 1.7;
Matches 83; Conservative 51; Mismatches 153; Indels 103; Gaps 20;
QY 22 EYFNINDSKILSLQNRKNTLVDTSGYNAEVSSEGDVQLNPFFDFKLGGSGEDRGKVI 81
DB 129 EGFNLYHFNDLTQLDKH-----DVSGYK-----FNLESIASINFTSGTTPQKAVPQ 175
QY 82 VTQENIVYNSWYSEFSISFWIRINKWVSNLPGYTIIDSVKNNWSGI---GIISNLFV 138
DB 176 TFNNHLASAKGCKQSLGFE---QNTVLSVLPIYHI-----SGLSVILRAVIEGTVR 225
QY 139 TLK--QNEDESEGISNFSIDISNNA--PCYNKWFVTVTNMMGNKVIYING-----KLID 189
DB 226 LVYKKTQDMLTQIK-TYPITHMSLVLPOTLAWLMDAGLTQPSLEKILLGGAKLSPOLIE 284
QY 190 ---TIKVKELTGINFESKITTEINKIP-----DTGLITSDSDNINNMWIRDFYIFAK-E 238
DB 285 QALTYELPVNSFGWETFCQFLASPQMLKERFDT--VGRKPSENEVEVKIKNPWAYGHGE 342
QY 239 LQGDINILFNSLOITNVYKDYWGNDLRYNKEYVMNIDYLNRYWYANSROIVENTRNN 298
DB 343 LLIKGENVM-NGLYLPKYLKDTFND-GYFQTGDAEID----- 379
QY 299 NDFNEGYKIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMADNHSSTEDYI 358
DB 380 ---DEGYVIYDR---RKDLISGENIY-----PYOI---ETIAKDFEGIEDAV 420
QY 359 AIGLRBQT-----KDINDNIIFQ 376
DB 421 CVGISDDTWGQVPIIYVTVNODINOTELIE 450
RESULT 13
US-10-092-411a-4341
; Sequence 4341, Application US/10092411a
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411a
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 4341
; LENGTH: 695
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-092-411a-4341
```

355	Ddb	355	NVRLKVKQFPLPMEINERDFKRIQOLHGDGVRYYSFIDTIGVNSGSHLYKYSQVKNKXNG	414
144	QY	144	EDSEQSI-----NFSYDISNAPGYNKWFVTVTNMGNMKIYINOKLIDTKVKELT	197
415	Ddb	415	KRFVNTRIENNGNFAAIGONELTYK---VTLPENFE----YVD-----NSTEVSFVN	461
198	QY	198	GINFESKIT-PEINKIPDTGLITSDSDNINMIWROFYIFAKELDGKDINILFNSLQYTNV	256
462	Ddb	462	GNVPSNVNPFVNFDFRONHTLTFSSGNLIG-RSAQDVARELPNKLINI-----	510
257	QY	257	VKDVGWMDLRYNKEYVMVNTDYLNRMYANSQIVFVNTRRNNDNFGYKILIKRIGNT	316
511	Ddb	511	-----RY--KLRPVNT-----STPREVTFNKAIKYTFSEYII-----NT	543
317	QY	317	NDTRVRGBDILYFDMTINNK-----AYNLFMK-----N	344
544	Ddb	544	NDNVTGCOOPFSINVMKNDDLSEQVKNKDIPSNTYLASYNKYNKLKERAQTVLDEETN	603
345	QY	345	ETMTADNHSTEDIYAIGLRQQT-----KDINDNIIFQIOPMNTYYVASOI	390
604	Ddb	604	NTPFNORYSQIDDLLEHLOTLINRVSAIREIND---KAQEMTDVAYDSTEL	654
RESULT 15				
US-10-092-411A-4465				
; Sequence 4465, Application US/10092411A				
; GENERAL INFORMATION:				
; APPLICANT: Lynn Boucette-Stamm et al				
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC				
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS				
; FILE REFERENCE: 032796-101				
; CURRENT APPLICATION NUMBER: US/10/092,411A				
; CURRENT FILING DATE: 2002-03-07				
; PRIOR APPLICATION NUMBER: US 09/134,001				
; PRIOR FILING DATE: 1998-08-13				
; PRIOR APPLICATION NUMBER: US 60/064,964				
; PRIOR FILING DATE: 1997-11-08				
; PRIOR APPLICATION NUMBER: US 60/055,779				
; PRIOR FILING DATE: 1997-08-14				
; NUMBER OF SEQ ID NOS: 5676				
; SEQ ID NO 4465				
; LENGTH: 527				
; TYPE: PRT				
; ORGANISM: Staphylococcus epidermidis				
US-10-092-411A-4465				
Query Match 4.2%; Score 100; DB 6; Length 527;				
Best Local Similarity 20.1%; Pred. No. 3.7;				
Matches 93; Conservative 73; Mismatches 187; Indels 110; Gaps 22;				
QY	4	PNTFSPYTNNSLLXDI---INEVFNNINDSKILSLQNRKNTLVDTSGYNAEV---SEBGD	57	
Ddb	117	PFQIOAITGDTNFSKIHFSOEGYLMWVEDYKSTIQRR--FVFDGRGFISAVRTIPDGD	174	
QY	58	VOLNPIFPDFDKLGSS--GEDRGKVIYVTONEN-----IVNYSMESFSISFWIRINK	108	
Ddb	175	NNKHYPFSKO---GEEIFVEDLNVTVTINKFQSRFKRVYTSSMAELIEERFQSYVERE	231	
QY	109	VSNLPGYFIIDSVNNGNSGSIILSNELVTL--KONEDSEQSINFSYDISNAPGYNKW	166	
Ddb	232	LNEDDSVIVASDERHNSMARTIDASSLCISITERKNKVYTOQL---YDSISRA-----	282	
QY	167	PFVIVTNMGMNKIYINGKLIDTPRKVELTGINFKSIITFEINKIPDPTGLITSDDNIN	226	
Ddb	283	YYCLVDYTOQANQM-----IEHYAGLNMDINLLRVTPFDFAKSLPNQSSOLYDT-YIG	333	
QY	227	MKIRDFYIFAKELGKD-----INTLNSLOYTNVVKYDGNDLRYN-KYYVMVNDY	278	
Ddb	334	LMI-----DOLDEIREIVNSLFOYIOGHKGQYKILKILTKSRONLENLIDEVAH	383	
QY	279	LNRYMYANSRQI-----VFNTRRNND-----FNWGYKILIKRIGNTNDRVRGGDI	326	

```

Db 384 LNDLYHQEKKEISDVIEDVIONKRETIIDETVPFEEDLYSVISKLR----- 430
QY 327 LYFDMTINNKAYNLFM-----KNETWYADNHSTEDIYAGLREOTKDINDNI 373
Db 431 VVVDLSLEPK---LFLQICCGAGIPQINKKRTDYVKH--MHNGYIIDDISQTVESLDYF 485
QY 374 IFQIQPMNNTYYASQIFKSNFNGENI-----SGICSIGT 408
Db 486 LAHLKNWNSYAYSML-TDDFSSINIHQINQLFKGDVSSGT 527

```

Search completed: November 7, 2002, 14:53:14  
 Job time : 20 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

CM protein - nucleic search, using frame\_plus\_p2n model  
Run on: November 7, 2002, 16:32:06 ; Search time 2721 Seconds  
(without alignments)  
4813.033 Million cell updates/sec

Title: US-09-910-186A-10  
Perfect score: 2391  
Sequence: 1 MTIPFIFSYNNLLKDI.....NYASLLESTHNGFVPVSE 450

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame\_p2n.model -DEV=xlh  
DB=/cgn2\_1/USPTO.spool/US09910186/runat\_04112002\_111617\_20836/app\_query.fasta\_1.647  
DB-GenEmbl -QFAST=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
LOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
USO=US09910186@cgn\_1\_1\_1616\_runat\_04112002\_111617\_20836 -NCPU=6 -ICPU=3  
NOR.XLPXY -NO.MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-XGAPOP=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:  
1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sv:\*  
13: gb\_un:\*  
14: gb\_yi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rnd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2386	99.8	4479	7	CBCPHGCL X53751 Clostridium
2	2386	99.8	4592	7	CSTCITOX D90210 Bacterioph
3	2386	99.8	4712	1	CBPHCNIM X71126 C.botulinum
4	2386	99.8	9613	1	CBCTOX X62389 Botulinum b
5	2386	99.8	9689	1	CBP1CMTCL X66433 Clostridium
6	2386	99.8	11747	1	AB061780 Clostridi
7	2386	99.8	12297	7	CSCBONT X72793 Clostridium
8	1824	76.3	11631	1	AB037920 Clostridi
9	1823	76.2	4049	1	CLOTNCO D38442 Clostridium
10	882.5	36.9	4043	1	CLOBN D49440 Clostridium
11	882.5	36.9	11614	1	AB037166 Clostridi
12	867.5	36.3	3952	1	CENTVD X54254 Clostridium
13	867.5	36.3	4087	7	S49407 S49407 type D neur
14	867.5	36.3	11584	7	AB012112 Clostridi
15	610	25.5	3876	1	AF300465 Clostridi
16	610	25.5	3876	1	AF300466 Clostridi
17	604.5	25.3	3937	1	CBONTG X74162 C.botulinum
18	604	25.3	11170	1	CBY13630 X13630 Clostridium
19	602.5	25.2	3759	1	AB088207 Clostridi
20	602.5	25.2	3838	1	AB039264 Clostridi
21	602	25.2	3876	1	AF295926 Clostridi
22	594.5	24.9	3759	1	AB082519 Clostridi
23	594.5	24.9	4030	1	CENTPE X52683 C.botulinum
24	594	24.8	3835	1	AB037704 Clostridi
25	594	24.8	3835	1	AB037705 Clostridi
26	594	24.8	3835	1	AB037706 Clostridi
27	594	24.8	3835	1	AB037707 Clostridi
28	594	24.8	3835	1	AB037708 Clostridi
29	594	24.8	3835	1	AB037709 Clostridi
30	594	24.8	3835	1	AB037710 Clostridi
31	594	24.8	3835	1	AB037711 Clostridi
32	594	24.8	3835	1	AB037712 Clostridi
33	594	24.8	3835	1	AB037713 Clostridi
34	594	24.8	3835	1	AB037714 Clostridi
35	594	24.8	3949	1	CBNTPE X52088 C.butyrum
36	593	24.8	4051	1	CBONTB X71343 C.botulinum
37	586.5	24.5	3891	6	AR000031 Sequence
38	586.5	24.5	3891	6	ARI69142 Sequence
39	586.5	24.5	3891	6	AX036248 Sequence
40	586.5	24.5	4292	1	CBOTAG X52066 Clostridium
41	586.5	24.5	4835	1	CLONEUR M30196 C.botulinum
42	586.5	24.5	15039	1	AF461540 Clostridi
43	583	24.4	4017	1	CSENEUTOXE X52089 C.botulinum
44	583	24.4	4017	6	AX088262 Sequence
45	581	24.3	4041	1	CLOBOTB M61186 Clostridium

ALIGNMENTS

RESULT 1

CBCPHGC1	CBCPHGC1	DB:	7	Gaps:	0
LOCUS	4479 bp DNA linear PHG 12-SEP-1993				
DEFINITION	Clostridium botulinum C phage gene for C1 neurotoxin.				
ACCESSION	X53751				
VERSION	X53751.1 GI:14905				
KEYWORDS	C1 neurotoxin; neurotoxin; toxin.				
SOURCE	Clostridium botulinum C phage.				
ORGANISM	Clostridium botulinum C phage				
REFERENCE	1 (bases 1 to 4479)				
AUTHORS	Popoff M.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-JUL-1990) Popoff M.R., Institut Pasteur, Anaerobies,				
REMARK	25 rue du Docteur Roux, 75724 Paris Cedex 15, France				
AUTHORS	revised by [3]				
TITLE	2 (bases 1 to 4479)				
JOURNAL	Hausler D., Eklund M.W., Kurazono H., Binz T., Niemann H.,				
REMARK	Gill D.M., Boquet P., and Popoff M.R.				
AUTHORS	Nucleotide sequence of Clostridium botulinum C1 neurotoxin				
TITLE	Nucleic Acids Res. 18 (16), 4924 (1990)				
JOURNAL	90370487				
REMARK	2204031				
AUTHORS	revised by [3]				
TITLE	3 (bases 1 to 4479)				
JOURNAL	Popoff M.R.				
REMARK	Direct Submission				
AUTHORS	Submitted (24-SEP-1990)				
TITLE	Location/Qualifiers				
JOURNAL	1. 4479				
FEATURES	/organism="Clostridium botulinum C phage"				
source	/strain="phage from 468C botulinum C. strain"				
	/db_xref="taxon:12448"				
	198..203				
	/note="pot. ribosome binding site"				
	214..4089				
	/note="C1 neurotoxin (AA 1-1291)"				
	/codon_start=1				
	/transl_table=1				
	/protein_id="CAA37780.1"				
	/db_xref="GI:14906"				
	/db_xref="SWISS-PROT:P18640"				
	/translation="MPITINNFNSDVPDKNKILYLDTHLNLANPEKAFRTIGNIW				
	YIPRFSNSPNLKPRTVSPKSGYIDNPSTDDKDPFLKEIKLFRKINSREI				
	GEELYRUSTDIPFGNNPTIPNFDVDFNSVDVKTROGNVNVKTSINPSVILTG				
	PREMIDPSTFKLNTNPAQGFALSIISTSPFMTYSNATNDVGEGRFSKE				
	KCMPIILLHMLHMLNLYGLAIPNDQISSVTSNIFSOYNKLEAYEAFEGP				
	TIDLPSARAFBEKALDYRSIAKRLNLTITANPSSFNKYLGEYKQLIRKRVV				
	ESSGEVTVNRKFLVELYNELTOIFENYAKIYNQNRKIYLSNVTPVTIANILDNV				
	YDIQGNFIPKSNLNVLFMGOQLSRNPAKRVNPNLLYLFKECHKAIDGRSLNKT				
	LDRELLVNTDLPFGIDISDKTDIFLRKDINEETEVYIPDNVSDQVILSKNTE				
	HGQDLVPSIDSESEILPGENQVYDNQNTQNDVLSNYYLESQKLSQNVDEFTFR				
	STEEALDMSKVTYFPPLANKVAGVQGLFLMWANDVDEFTNLRKDTLKDLS				
	VSAITPYTCRALNLSNVRGNEFAFVGVTVILLAPFTIPALGAPVLYSKVQE				
	RNEIKTLNCLGRIKRWKDSYEMMGWTLRLITOTFNLSIQMDSLSNYQAGAKA				
	KIDELLYKSGDKNKSNQVENLKSVDKISEAMNKNKIRFESGVILFKMLPK				
	VIDELNEPDRNTKAKLLINDSHNILLVGEVDKLLKAKVNSFQNTIPFNLFSNLS				
	LKDIENFENINBSKILSLQNRKNTLFTVSGYNAEVSDEGQVLPNIPFPDKLGS				
	GEDRKVTITONENIKLSVSEFISFWIRKNKWSNLPCTIIDSVKNNWSGIGI				
	ISNFTVPLTKONESEQISNFYSINNAPGYNKFPVTVTNNMGKMYINGKLD				
	TIKVELGIFNFTSIPFELNIPDTGLIISDSNINMLRDYIPAKELDGKDIL				
	FNSLOYTVNVDYDGNLRYNKEYMNYNIDLYNRYKANSRQIVFTRRNNDPNEG				
	KIIIRKRNNTNRGGLDLYFDMTINNKATNLMFKNETMTADHSTEDYVAGIR				
	EQTKDINDNIIFQIPQNNYTYASQIFKSNFNGENISGICISGTYRFLRGGDWYRHN				
	YLVPTVKQGNVYASLESSTHWGFVPVSE"				
BASE COUNT	1817 a 455 c 693 g 1514 t.				
ORIGIN					
Alignment Scores:					
Pred. No.:	3,4e-152	Length:	4479		
Score:	2386.00	Matches:	449		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	99.79%	Indels:	0		



```

Db 3760 ATGAAGAATGAACATGATGATCCAGATATCATAGTACTGAGATATATATGCTATAGGT 3819
QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
Db 3820 TTAAGAGAACAAACAAAGATATAATGATATATATATATTTCAAAATACAAACCAATGAAT 3879
QY 382 AsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401
Db 3880 AATACATATATATACCATCTCAATATATTAATCAATTTAATGAGGAGAAATATTTCT 3939
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsn 421
Db 3940 GGAATATGTTCAATAGTACTTATGCTTTTACACTTGGAGTGATGTATAGACACAAT 3999
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
Db 4000 TATTTGGTCCCTACTGTGAAGCARGAATATATGCTTATATTAGATCAACATCAACT 4059
QY 442 HisTrpGlyPheValProValSerGlu 450
Db 4060 CATTGGGGTTTGTACCTCTGAAGTGA 4086

RESULT 2
CSTCITOX 4592 bp DNA linear PHG 23-JUN-1999
DEFINITION Bacteriophage c-st (from C. botulinum) C1-tox gene for botulinum C1
neurotoxin.
ACCESSION D90210
VERSION D90210.1 GI:217780
KEYWORDS botulinum toxin; neurotoxin.
SOURCE Bacteriophage c-st (from C. botulinum type C-Stockholm) DNA.
ORGANISM
VIRUSES.
REFERENCE
1 (bases 1 to 4592)
Kimura, K., Fujii, N., Tsuzuki, K., Murakami, T., Indoh, T.,
Yokosawa, N., Takeshi, K., Syuto, B. and Oguma, K.
The complete nucleotide sequence of the gene coding for botulinum
type C1 toxin in the C-St phage genome
JOURNAL Biochem. Biophys. Res. Commun. 171 (3), 1304-1311 (1990)
MEDLINE 9102498
REFERENCE
2 (bases 36 to 2143)
Kimura, K., Fujii, N., Tsuzuki, K., Murakami, T., Indoh, T., Yokosawa, N.
and Oguma, K.
Cloning of the structural gene for Clostridium botulinum type C1
toxin and whole nucleotide sequence of its light chain component
JOURNAL Appl. Environ. Microbiol. 57 (4), 1168-1172 (1991)
MEDLINE 91262468
COMMENT These data kindly submitted in computer readable form by: Kouichi
Kimura
Department of Microbiology
Sapporo Medical College
1 South, 17 West
Sapporo 060
Japan
E-mail: nfujii@niguts.nig.ac.jp@vax2.nlm.nih.gov
Phone: 011-611-2111 x2334
Fax: 011-612-5861

FEATURES
source
Location/Qualifiers
1..4592
/organism="Bacteriophage c-st"
/db_xref="taxon:12336"
-35_signal
255..260
-10_signal
293..298
RBS
305..310
/note="SP sequence"
321..4196
/note="ORF for C1 toxin"
/codon_start=1
/transl_table=11
/protein_id="BAA14235.1"
/db_xref="GI:217781"
/translation="MPIINNNYSDPVDNKNILYLDTHLNLANEPEKAFRTGNW

```

Copied from 10910186 on 05-05-2004

```

VIPDRESNPNLNKPRVTSPKSGYVDPNLSYSDSDKDFLKEIKLFKINSREI
GEELIYRLSTDIPEPGANNPTINTDFDVFNSVDVKTROGNNWKTGINTFSVITG
PREMIIDPSTFKLTNNTPAAQEGFALSTISISPEFMLTYSNATNDGEGRESKSE
FCMDPILILHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLHMLH
TIDIPISARKYFEKALDIYRSIAKRLNSITANSPFNKIIGERKQLIRIYVW
ESSEGVTVNRNFVLYNELTOFTFENYAKIYNQNRKLYLSNVYTPVPTANLDDNV
YDQNGFNIPKSNLFLMGONLSRNPALRKVPNNMLYLFTEFKHAKIDGRSLYKNT
LDCLRLVKTDLDFIGDISDKIDIFLRKIDNEETEVYYPDNVSVDQVILSKNTSE
HGDDLLVPSIDSESEILPGENOVFYDNRQNTQVDYLSYVYLSQKLSNVEDFTFR
SIEBALONSACVYTFPLANKYNAGVGGFLMWANDVDEDTNLRKDDTKIDKISD
VSAIIPICPALNISNSYRGNTFEATVATGVTILLLEAPFETIPALGARVIVYKQE
RNEIKITIDNLEORIKRWKDSYEMMGITWLSRIITQFNLSIQMSIDSLNQAQALKA
KIDLEYKISGSKDENIKSQVENLKNLSLDVKISEAMNNKINFKIREGVSYVFLKMLPK
VIDELNEFRNTAKLNLIDSHNIIIVGEVDKLAKKVNNSFQNTPIFNISFNKLSL
LKDINEYFNNDKILSLQNRKNTLVDTSYNAEVEEGDQVLPAPFPDFKLGSS
GEDRGKIVITONENIYVNSYSEFISFWIRINKWSNLPGYTIIDSVKNNSSWSGIC
ISNELVETLQKNEDESEINSEYDINSNAPYKKNKPEVTVNNMKMNTIYINGKID
TNSVELIGINFSTIIFENKIPDGTGLIISDSNINMWIRDEYIFAKELDGKIDJIL
FNSQYTNVVKDYNGLRINKYIMVNDILARYMYANSRQIVFVIRNNNDPNEGY
KIIIRKINGTNDTRVGGDILYFDMTINKKYNLFNMKNMETIADHNSHEDIIAIGLR
BQTKIDNDNIIFQIPMNNTYVYASQIFKSNFNGENISGICSGTCTYFRILGDDWYRHN
YLVPTVKGNYVASLLESTTHMGFPVPE"
BASE COUNT 1869 a 460 c 714 g 1549 t
ORIGIN
Alignment Scores:
Pred. No.: 3,48e-152 Length: 4592
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 7 Gaps: 0
US-09-910-186a-10 (1-450) x CSTCITOX (1-4592)
QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 2847 ACATACCCCTTAATATATTTTTCATATACTATAATATCTTTTAAAGATATAATTAAT 2906
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2907 GAATATTTCAATAATATTAATGATTCACAAATAATTTTGACCTCAAAACAGAAAATACT 2966
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 2967 TTACTGGATACATCAGCATATATCGCAGAGTGAGTGAGGAGGGGATGTTACGCTTAAT 3026
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 3027 CCAATATTTCCATTTGACTTTAAATAGGTAGTTCAGGGGAGGATAGAGTAAAGTTATA 3086
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 3087 GTAACCCAGAAATGAAATATATGATATATAATCTATGATGATGAAAGTTTATGATAGTTT 3146
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121
Db 3147 TGCATTAGATAAATAAATGAGTAAAGTAAATTTTACCTGGATATATAATGATAGTCTCT 3206
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 3207 AAAAATAACTCAGGTGGAGTAGTAGTATTATAGTAATTTTATAGTATTTTACTTTTAAAA 3266
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161
Db 3267 CAAAATGAAGATAGTGAACAAAAGTATAAATTTTAGTTATGATATATCAAAATAGTCTCT 3326
QY 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3327 GGATACAATAAATGCGTTTTTGTAACTGTTACTACAATATGATGGGAATATGAGATT 3386
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201

```



3873 TATATAAATGGAATAATTAATGATCTATAAAGTTAAAGAACTAACTGGAATTAATTT 3932  
 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221  
 3933 ACAGAACTATACATTTGAAATAATAAATTAATTCAGATACCGTTTGATCTACATTCAGAT 3992  
 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrllePheAlaLysGluLeuAspGly 241  
 3993 TCTGATAACATCAATATGCTGATAAGAGATTTTATATATTTGCTAAAGAAATAGATGGT 4052  
 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnThrhrAsnValValLysAspTyrlle 261  
 4053 AAGATATTAATATATTTAATAGCTGCAATATCTAATATGTTGTAAGAGATTAATGG 4112  
 262 GlyAsnAspLeuArgTyrlleAsnLysGluTyrlleMetValAsnIleAspTyrlleAsnArg 281  
 4113 GCAATGATTTAAGATATAATAAAGAAATATTATATGTTAATATAGATTTAATAATAGA 4172  
 282 TyrMetTyrlleAsnSerArgGlnIleValPheAsnThrArgAsnAsnAspPhe 301  
 4173 TATATGATGCAACTCAGCAAAATGTTTTTAAATACAGTAAAGATAAATAGATTC 4232  
 302 AsnGluGlyTyrlleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
 4233 AATGAAGATATAAATATAATAAAGAAATCAGAGAAATCAAAATGATCTAGATGA 4292  
 322 ArgGlyGlyAspIleLeuTyrllePheAspMetThrIleAsnAsnLysAlaTyrlleAsnLeuPhe 341  
 4293 CGAGAGGAGATATTTATATTTGATGATGCAATTAATAACAAGATATAATTTGTTT 4352  
 342 MetLysAsnGluThrMetTyrlleAlaAspAsnHisSerThrGluAspIleTyrlleAlaIleGly 361  
 4353 ATGAAGATGAACATATGATGCAGATATCATAGTACTGAGATATATATGCTATAGGT 4412  
 362 LeuArgGluThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
 4413 TTAAGAGAACAAACAAGATATAAATGATATATATATTTCAATACACCAATGAAT 4472  
 382 AsnThrTyrlleTyrlleAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
 4473 AATCTATATATAGCATCTCAATATTTAAATCAAAATTTATGAGAGAAATATTTCT 4532  
 402 GlyIleCysSerIleGlyThrTyrlleArgPheArgLeuGlyGlyAspTyrlleArgHisAsn 421  
 4533 GGAATATGTTCAATAGGATCTATCGTTTATGACTTGAGAGTGGAGGTGATGATAGACAAAT 4592  
 422 TyrLeuValProThrValLysGlnGlyAsnTyrlleAlaSerLeuGluSerThrSerThr 441  
 4593 TATTGGTCCCTACTGTGAGAGAGAAATTAATGCTTCTATTAGATCAACATCACT 4652  
 442 HisTrpGlyPheValProValSerGlu 450  
 4653 CATTGGGGTTTGTACCTGTAAGTAA 4679

RESULT 4  
 LOCUS CBCTOX 9613 bp DNA linear BCT 29-JUN-1999  
 DEFINITION Botulinum bacteriophage genes for HA-17, HA-33, nontoxic components and C1 toxin.  
 ACCESSION X62389  
 VERSION X62389.1 GI:558175  
 KEYWORDS botulinum toxin; C1 toxin; HA-17 gene; HA-33 gene; Hemagglutinin; neurotoxin; nontoxic; toxin.  
 SOURCE Clostridium botulinum.  
 ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.  
 REFERENCE 1 (bases 1 to 9613)  
 AUTHORS Kimura, K., Fujii, N., Tsuzuki, K., Murakami, T., Indoh, T., Yokosawa, N., Takeshi, K., Syuto, B. and Oguma, K.  
 TITLE The complete nucleotide sequence of the gene coding for botulinum type C1 toxin in the C-ST phage genome

JOURNAL Biochem. Biophys. Res. Commun. 171 (3), 1304-1311 (1990)  
 MEDLINE 91024998  
 PUBMED 222445  
 REFERENCE 2 (bases 1 to 9613)  
 AUTHORS Tsuzuki, K., Kimura, K., Fujii, N., Yokosawa, N. and Oguma, K.  
 TITLE Nucleotide sequence of the gene for one of the components of hemagglutinin produced by clostridium botulinum type C Unpublished  
 JOURNAL  
 REFERENCE 3 (bases 1 to 9613)  
 AUTHORS Tsuzuki, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-DEC-1991) K. Tsuzuki, Sapporo Medical College, Dept of Microbiology, South 1 West 17, Sapporo 060, JAPAN  
 COMMENT On Oct 15, 1994 this sequence version replaced gi:40386. See also X53041 & D90210.  
 FEATURES  
 source  
 1. 9613  
 /organism="Clostridium botulinum"  
 /strain="type C Stockholm, c-st"  
 /db\_xref="taxon:1491"  
 complement(111..551)  
 /gene="HA-17"  
 complement(111..551)  
 /gene="HA-17"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA44260.1"  
 /db\_xref="GI:40387"  
 /db\_xref="SWISS-PROT:P46083"  
 /translation="MSERTFLPENGKIKSFSDSLYITSSGALSFNTSSLNQK WKLEYISSNGFRFSVAEPKYLAYNDYGYLYSSSSNNLWNPKIAINSYIICL SIVNVDYAWTIYDNNNNITDQPIILNPLNPDINNNOILKLEKL"  
 complement(560..565)  
 complement(614..1474)  
 /gene="HA-33"  
 complement(614..1474)  
 /gene="HA-33"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA44261.1"  
 /db\_xref="GI:40388"  
 /db\_xref="SWISS-PROT:P46084"  
 /translation="MSQTNANDLRNNEVFIFSPSNNNTKVLKISQSEVKLNKLSGA NQKRLIYDNTKQAYIKYKMDNTSLITLWNPASSVSKYDINGDNQYLLQNTLSR NVITRYNTNPNLVQYNDITLMTSTSSNQFFKSNCIYEALNNRNTKLTQLNS DRTSKNLNQIIVLWQFDSRQKWTIEYNETKSATLCKQENNYLWINSNNYV EYVOSTSLIQYWNINYLNDNASKYILYLNQDTRVLDVNSQIANGTHVIVDSYHGN TNOQWINTLT"  
 complement(1482..1486)  
 1726..1731  
 1737..5327  
 /gene="nontoxic"  
 1737..5327  
 /gene="nontoxic"  
 /codon\_start=1  
 /transl\_table=11  
 /product="non-toxic components"  
 /protein\_id="CAA44262.1"  
 /db\_xref="GI:40389"  
 /db\_xref="SWISS-PROT:P46081"  
 /translation="MDINDLNTNSPYDNKNVIVRARKNTTFFKAFKAPNTWAPE RYRVEDIAEYKLDGGIYDSNLSQDSERENFLQAIITLLKRNNTISGKQLLSLI STAIPFVGIVGGYSSNPITFGTKPKSNKLSLVTSTIPFPGGYRETNYEQNI NKFEASNIIFGPGSNIVNNVYKKNDAENGMTAGIVQPLLTKYKNTYIDP AMELTKLISLFLYGIKESDNLVYPRLEGTNDKQFSQNLIDLLISGVDLFEI NTNYWFTNSYFPNSIKMEKYNIKYTEEGNNAIGNDKLKKQKQFQINVDIWNL NLNTFCQFSNIPDRESNALKHFRKYQYMDTYDNYINGNQITNKTPLSDNNK TNIISKPKVNNLVNENNIILKSNYDGLKGTEDFYSTYKIPNEYEYRFDNSD TNNFNLSIEVDSIPETIINPYKNSDNLVETQITSEVTTTALSINYLQAI NFPLNNTISIEVDSKSVSSKDLVYSLFDNLNSLYETIKNDGPIDTKYLYLKEVF TNNFNLSIEVDSKSVSSKDLVYSLFDNLNSLYETIKNDGPIDTKYLYLKEVF KNYSPDNLQEDSMQGINVYLPKALNLTNSFVEEYQDSGLISLISKKNL REPIEDIDSLGLSFKDLANKLYISNIVPKIYFSLQWHTWEYYSQYFE LKIMAKSILQAQESLVKQIVONKTLDSKASIPDILKLTRETEKIFIDLSNESQIS

Copied from 10910186 on 05-05-2004

MNRVDFNLKASICVVEDIYPKFSIYMEKYINNINIKITREFIOQRTNINDNEKSILL  
 NSVTEKIDPFKDIQSIKNFNNSOEOVQWKEILSPYQILLFASKGPNINIEISGK  
 NTLQVTESELVYGVNGESLYLKSNEIKFSNFFNGLTNFTNFIQWFRGKND  
 KDLGLGNKNNCGWIEYFEDNGLVEIIDSNGQESVLYSNIINDWYIISISVRL  
 KDQLLFINDKNNYNSIQIUSIYSTNIIISLVNKNNSIYVEELSVLNDPITSEVIR  
 NYFSYLDNSYIRDSLSLEYKNKYOLNYVPEPSTLYEVNDNNKSYLSKNTGDINI  
 SVYFLKLNIDESKVTYQKWDICIIICVLDGTEKYLIDISPNRNRIOLVSKNOKKITV  
 NTLDFRPDCITFSYNDKYFSLSLRDGYNMNICDNNKVPKGHAHLWLES"  
 RBS 5330..5335  
 gene 5342..9217  
 /gene="Cl toxin"  
 CDS 5342..9217  
 /gene="Cl toxin"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA44263.1"  
 /db\_xref="GI:40390"  
 /db\_xref="SWISS-PROT:P18640"  
 /translation="MPTINNENISDPVKNKMLILDLNLTLANEPKAFRIQGNW  
 VIPDRSRNPNLNKPPRTGKSGYDNYLSDSKDQDFLKEIKLFRINSREI  
 GEELIYRLSDIDPPGNNTPIINTEFDYDFNSVDVKTQGNWVKTGINPSVITG  
 PRENIIDPSTFKLNTNTFAOEGGALSIIISIPMLIYSNATNDWGBGRFSKE  
 FCMDFILMLHMLHAMNLNLYGAIIPNDOTISSVTNFIYSOYNVLEAYEIAFGGP  
 TIDLIPKSAKYFEERALDYYSIAKRLANSITANPSSFNKYIGEVKOKLIRKRFVY  
 ESSEGVYNRKVELYNELQIFTEFNKAIYVQNRKILYSNVIPTANILDDNV  
 IYDIONGFIPKSNLWLFQNLSPALRKVPNPNMILYLFKFKCHKAIDGRSLYNT  
 LDCRELKNTDLPFDIDSDVKTDFIFLRKDINEETEVIIYPDNYSVDQVILSKNTSE  
 HGOLDLYPDISSESILPQENQVYDNRQNDVINSYIYLESOKLSDNVEDFTFR  
 SIEBALONSAKYVYPTLANKVNGVQGLFLWANDVDEFTNLRKUTLOKISD  
 VSAIIPYIGPALNISVRRGNFTAFAGVTIILEAPPEPTIIPALGAFVYSKVOE  
 RNEIITDNCRLORIKRWKDSYEWMMGTLSRIITOPNNISYOMYSVLYOAGNKA  
 KIDLEKYSKGSKKNIKSOVENLKNLSLVKISEAMNINKFIRECSVTYLFKNMPLK  
 VIDELNDFDNRKILNIIILGVEDKLRKAKVNNSPQNTIPNIFISYNNSL  
 LKDIINEFYNNIDSKILSLONKNTLDVTGYNAEVSEGVQVLPNIPFDFKIGSS  
 GEDRGKVIYTONENIYNSWFSFISIPRIKWNKVNLPGYTIIDSVKNSGWSIGI  
 ISNLFVTLKONDSQSINFSDVSNAPGNKWFVTVTNMNMKNKIYINGKLID  
 TIKVKELTGINFSEKTIIFELNIPDGLITSDSDNINMWIRDFYIFAKELDKGNIL  
 FNSLOVNVVYKWDNDLYNKEYYAVNDYLNRYKANSROIVFTRPNNDNPEGY  
 KLIIRKRGWNTNTRVGGDILYFDMTINKKYNLFEMKNEMTADNHSTEDIIAIGLR  
 EQKLDNDNIIFIQPMNNIYIYASQIFKSNFNGENISGICSIGITRPRFGDWTNRH  
 YLVPYKQGNYSLLSESTHMGFVPEV"  
 BASE COUNT 3846 a 1003 c 1348 g 3416 t  
 ORIGIN

## Alignment Scores:

Aligned. No.: 7 23e-152 Length: 9613  
 Score: 2386.00 Matches: 449  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Every Match: 99.79% Indels: 0  
 Gaps: 1

us-09-910-186a-10 (1-450) x CBCTOX (1-9613)

Qy	2	ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn	21
Db	7868	ACAAATACCTTTAATATTTTTCATATACTAATAATCTTTTATTAAGAGATAATAAT	7927
Qy	22	GlutYrPheAsnAsnIleAsnAspSerIySleuSerLeuGlnAsnArgLysAsnThr	41
Db	7928	GAATATTTCAAT	7987
Qy	42	LeuValAspThrSerGlyTyThrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn	61
Db	7988	TTAGTGGATACATCAGGATATAATCCAGAGTCTAGTCAAGAGGCGATGTCAGCTAAT	8047
Qy	62	ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle	81
Db	8048	CCAAATATTTCCATTTGACTTTTAAATAGGTAGTTTCAGGGGAGGATACAGGATAAGTATA	8107
Qy	82	ValThrGlnAsnGluAsnIleValTyThrAsnSerMetTyThrGluSerPheSerIleSerPhe	101
Db	8108	GTAACCCAGGAATGAAATATATCTATATAATCTATGATCAAGAGTTTATAGCATTTACTTTT	8167
Qy	102	TrpIleArgIleAsnLysTyrtpValSerAsnLeuProGlyTyThrIleIleAspSerVal	121
Db	8168	TGGATTAGAATAAATAAATGGTAAGTAATTTACCTGGATATATAATAATGATGTGT	8227
Qy	122	LysAsnAsnSerGlyTyrtpSerIleGlyIleSerAsnPheLeuValPheThrLeuLys	141
Db	8228	AAAAATAACTCAGGTTCAGGTATAGGTATATTAGTAATTTTCTAGTATTACTTTAAAA	8287
Qy	142	GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyThrAspIleSerAsnAsnAlaPro	161
Db	8288	CAAAATGAAGATAGTAGCAAAAGTATAAATTTAGTATATGATATATCAAAATATGCTCT	8347
Qy	162	GlyTyThrAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle	181
Db	8348	GGATACAATAAATGGTTTTTGTAACTGTTTACTAACAAATATGATGGGAAATATGAAGATT	8407
Qy	182	TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe	201
Db	8408	TATATAATGGAATAATTAATAGATCTATATAAAGTTAAAGAACTAACTGGAATTAATTTT	8467
Qy	202	SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp	221
Db	8468	AGCAAACTATAACATTTGAAATAAATAAATTCAGATACCGGTTTGATTACTTTCAGAT	8527
Qy	222	SerAspAsnIleAsnMetTrpIleArgAspPheTyThrIlePheAlaLysGluLeuAspGly	241
Db	8528	TCTGATAACATCAATATGTCGATAAGAGATTTTATATATTTTCTAAAGAATATAGATGT	8587
Qy	242	LysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValValLysAspTyTrp	261
Db	8588	AAAGATATATATATATATATATATATAGCTTCCATATATCTAAAGTTCTAAAGATATTGG	8647
Qy	262	GlyAsnAspLeuArgTyThrAsnLysGluTyTyThrMetValAsnIleAspTyThrLeuAsnArg	281
Db	8648	GGAAATGATTTAAGATATAAAGATAATATATGTTTAAATATAGATATTATTTAAATAGA	8707
Qy	282	TyrMetTyThrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPhe	301
Db	8708	TATATGATGCGAACTCAGCAAAATTTGTTTAAATACACGTAGAAATATAAIGACTTC	8767
Qy	302	AsnGluGlyTyThrLysIleIleLysArgIleAArgLysAsnThrAsnAspThrArgVal	321
Db	8768	AATGAAGGATATAAATATATAAATAAAGAAATCAGAGGAAATACAAATGATACAGATTA	8827
Qy	322	ArgGlyGlyAspIleLeuTyThrPheAspMetThrIleAsnAsnLysAlaTyThrAsnLeuPhe	341
Db	8828	CGAGGAGGAGATATTTTATATTTTGTATGACCAATTAATAACAAAGCATATAATTTGTTT	8887
Qy	342	MetLysAsnGluThrMetTyThrAlaAspAsnHisSerThrGluAspIleTyThrAlaIleGly	361
Db	8888	ATGAAGATGAACACTATGATGCAGATAATCATAGTACTGAAGATATATATGCTATAGGT	8947
Qy	362	LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn	381
Db	8948	TTAAGAGAACAACAAGAGGATATAAATGATAATATATATATATTTCAAAATACAGCAATGAAT	9007
Qy	382	AsnThrTyThrTyThrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer	401
Db	9008	AATACTTATTTATACGCATCTCAAAATATTTAAATCAAAATTTAATGGAGAAATATTTCT	9067
Qy	402	GlyIleCysSerIleGlyTyThrArgPheArgGluGlyAspTyThrArgHisAsn	421
Db	9068	GGAATATGTTCAATAGGTACTTATCGTTTGTAGACTTGGAGGTGATTGGTATAGACAAAT	9127
Qy	422	TyrLeuValProThrValLysGlnGlyAsnTyThrAlaSerLeuLeuGluSerThrSerThr	441
Db	9128	TATTTGGTCCCTACTGTGAAGACAGGAAATATGCTTCATTTATAGAAATCAACATCACT	9187
Qy	442	HisTrpGlyPheValProValSerGlu	450
Db	9188	CATTGGGGTTTGTACCTGTAAGTGAA	9214

```

RESULT 5
CBP1CNTC1
LOCUS      Clostridium botulinum phage 1C, CHn-14, CHn-33, CHn-138 and Bont/C1
DEFINITION
ACCESSION  X66433.1 GI:509271
VERSION     haemagglutinin; neurotoxin.
KEYWORDS    Clostridium botulinum.
SOURCE      Clostridium botulinum
ORGANISM    Clostridium.
REFERENCE   1 (bases 1 to 9689)
AUTHORS     Hauser,D.F., Eklund,M.W. and Popoff,M.R.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 9689)
AUTHORS     Hauser,D.F.
DIRECT SUBMISSION
SUBMITTED   (25-MAY-1992) D.F. Hauser, Institut Pasteur, Unite des
JOURNAL     Toxines Microbiennes, Institut Pasteur, 28 rue duDocteur-Roux,
            F-75724 Paris Cedex, FRANCE
FEATURES    Location/Qualifiers
             1..9689
             /organism="Clostridium botulinum"
             /db_xref="taxon:1491"
             complement(253..647)
             /gene="CHn-14"
             complement(253..633)
             /gene="CHn-14"
             /codon_start=1
             /transl_table=11
             /product="14kDa protein associated with Bont
             /Cl-haemagglutinin complex"
             /protein_id="CAA47057.1"
             /db_xref="GI:509272"
             /db_xref="SPRAMEL:045915"
             /translation="MSSERTFLPNQYKIKLSFSDSLYTYSSGALSPENTSLLDNQK
             WKLEYISSNGFRFSNVAEPKRYLAYNDYGFYLSSSNNLSLWNPKIATINSYIICTL
             SIVNTDYATYIDNNNNIHDHFL"
             complement(642..647)
             /gene="CHn-14"
             complement(696..1556)
             /gene="CHn-33"
             complement(696..1556)
             /gene="CHn-33"
             /codon_start=1
             /transl_table=11
             /product="33kDa protein associated with Bont
             /Cl-haemagglutinin complex"
             /protein_id="CAA47058.1"
             /db_xref="GI:509273"
             /translation="MSTANDLRNNEVFFISPSNNTNKLVDKISQSEVKLNKLKSGA
             NQKRLIYDINKQAYKLYMDNTSLIITWNPALSSVSVKTDGNDQYWLQNTISR
             NVIIRYNNPLNQLYNDITLWVSTQSSNQFPKSCIYEALNRRNKQTLQNS
             DRFLSKNLQSIYLVQWQDSSRQKWIIEYNETKSAYTLKCOENRNYLTWQNSNYY
             EYQSTQSLQYNNIYLDNDASKYILYNQDTRVLDVYNSQLANGTHVIVDSYHGN
             TNOQWILNI"
             complement(1564..1568)
             /gene="CHn-33"
             complement(1564..1568)
             /gene="CHn-33"
             join(1803..1808,1819..5409)
             /gene="CHn-138"
             1803..1808
             /gene="CHn-138"
             1819..5409
             /gene="CHn-138"
             /codon_start=1
             /transl_table=11
             /product="138kDa protein associated with Bont
             /Cl-haemagglutinin complex"
             /protein_id="CAA47059.1"
             /db_xref="GI:4376225"

/translation="MDINDNLNINSPVNDKNVIVVARKTNTFFKAKVAPNIWVAPE
RYTEPELDAEETKLGGIYDSNFLQSDSERENFLQAILILKRNININISQKLLSLI
STAIPFYGGYSSPNIPTFGTKPKSNKLNLSVLTSTIPFPFGGRETNYIESQN
NKFYASNIIFGSGNIVNNVYIKKXDAENGMTMAEIVFYQVLLTYKYNKFIDP
AMELTCLIKSLYFLXGIFKPSNLYVYFRLTELDNKQFSQLNIDLLISGVDLEFI
NTNPFYNTSYFNSIKMPEKYNKIETEGNAIGNDKLRKQKQFQVQDQWNLN
NLAVFCOSFNSIIPDRFSNALKHFYKQVYTDYTDYNYNGFVNGQINTKPLSNKN
TNLISXPKVAVNLVNNISLMSKNIYDGLGKTDEYSTYKIPYNEEYREYRSD
NPNLNISIEVDSEIFEIIDIINPKNSNLVFTQILTSMTETHTALSNYLQAOI
TNNEFTLSDFSKVSSKDLVSLYFSLNDNLSYLETIKNDGPIDHAKIYILWLEKVF
KNYSFINUTQEIDSCGNEVYVWFGKALNINLNTSVSEVTEQDSGASLISKDLNL
REFNIEIDISLGLSFKDLNNKLEYISKNIVFYFKYFSDQWMTYSSYFFE
LICMAKOSILAQESLVKQVQKFTLSKASIPPTDKLIRETTEKTDLSNESQIS
MNPVDNFKASICVFEVDIYKFIYSNEKYNINNIKRETIQRTKTDINDNEKISLI
NSYFTKIDFKLDIOSIKNFNSQVQVKRILSPYQLLFSKGPSNIIEDISGK
NTLIQVTESELYGVNGESLYLKSNETIKFSNKEFTNGLTNNETICFWLRTGKND
DKTLGLGNKNNCGWEIIFEDNGLVFEIISDNGQESYLSNINDNWTYISISVDRL
KQDILLIFNDKNVANSIDPQISYSTNIISLVNKNNSIYVEELSVLNDNPNITSEVIR
NYFSYDINSYIRDSKSLLEYKNKYNLYNVPETSLYEVNDNNSYLSLNTKIDINI
SSYKFLINIDSEKTVQKWDECIICVLDGTEKYLDISPENNRIOQLSVSKNAKITV
NTDLRFPDCTESYNDKYFSLSLRDGDYNNWICNDNNKVPKGALHWILES"
5408..5413
/gene="Bont/C1"
5408..5413
/gene="Bont/C1"
5424..9299
/gene="Bont/C1"
5424..9299
/gene="Bont/C1"
5424..9299
/gene="Bont/C1"
/codon_start=1
/transl_table=11
/product="Botulin Cl neurotoxin"
/protein_id="CAA47060.1"
/db_xref="GI:509275"
/db_xref="SWISS-PROT:P18640"
/translation="MPTTNNRVSDPVNDKNILYLDTHLNLANEPEKAFRITGNIV
VIDRPSRNSPNLNPAPRVTSFKSYDPNVLSDSDKDTLKLILKFRINSREI
GEEILYRLSDIIPFGNNPTINFDVDFNSVVKTRQGNWVKTSGINSFVLTG
PRENIDIPETSTFKLNNTPAQEGFGALSIIISPRMLTYSNATNDVGEGRFSKE
FCMDPILILMHELNHAMNLYGAIAPNDQTISSVTSNIFYSQYNKLEAIYAFGGP
TIDLIPKARKYFEKALDYIYSIAKRLNSITANPSFNKYIGETQKILKRYFVY
ESSGEVTVNRNKFVELYNELTQIETFNATYQVNRKTYLSNVTPTVNTANLDQNV
YDQNGPNIKSNLNVLWGNLSRNPALKRVNENMLYLFKCHKAIDGRSLYKNT
LDGRELLVNTDLPFGDIDSDVKTDLFLKNDNEETVLYYPDNVSDQVILSKNTSE
HGQDLIPSDISEILFGENQVYDNRQTQVNLNSYXLESQKLSNDVEDFTFR
SIEALDNSAKVITYPTLANKVNAVGGGLFMANDVEFTTILNKLKDLKDISD
VSAIIPYIPALNISNRGNTEAFVGTVILEAPEETIPALGAFVLYSKVQE
RNEIITIDNLCLEQRKRWKSDYEMWMTLSRIITQFNISYQMTDSLYOAGAICA
KIDLEYKYSQSKENIKSQVENKNSLDVKISEAMNNIKFIRECSYTFKNNLPK
VIDELNEFDRTAKLINLIDSHNIIIVGEVDKAKAKVNSQNTIPFNIFYTNNSL
LKDIINEFYNNIDSKILSLQNRKNTLVDTSGVNAEVSQVQVLPNIPFPDFKLGSS
GEDRGKVIYQENIYVNSMYESFISFWIRINKVSNLPGVTIIDSVKNSGWSIGI
ISNPLVFTLQKQNSDSQSINFSYDINNAFGNNFFVYTNMGMKVIYNGKLIID
TIKVKELTGINFSTITFEINKIPDRTGLITSDDNINMWIRYIFAKELDQKDNIL
FNSQLYTNVYKWDNDRLYKNEYVMVNDIYNRYMANSROIVFNTRNNNDNEG
KIIIRIRGNTNDRVRGGDILYFDMTINNKAYLFMKNETYADNHSDEIYALGLR
EQKIDINDNTIIFOQPMNNTYYASQIFKSNFNGENISGICSTGYFRFLGQDWYRHN
YLYPTVKQGNVYASLLESTSHMGFVPSV"
9385..9424
/gene="Bont/C1"
9385..9424
/gene="Bont/C1"
/note="putative"

BASE COUNT 3865 a 1022 c 1359 g 3443 t
ORIGIN

Alignment Scores:
Pred. No.:      7,29e-152      Length:      9689
Score:          2386.00      Matches:      449
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      99.79%      Indels:      0
Db:              1      Gaps:      0

```

```

US-09-910-186a-10 (1-450) x CBPICNTC1 (1-9689)
QY 2 ThrileProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleAsn 21
Db 7950 ACAATACCTTTAATATTTTTCATATACTAATAATCTTTATTAAGAGATATAAT 8009
QY 22 GluTyPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 8010 GAATATTCAATAATTAATTAATGATTCAAAATTTTGAGCTACAAAACAGAAAATACT 8069
QY 42 LeuValAspThrSerGlyTyAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 8070 TTAGTGGATACATCAGGATATAATGCAGAGTGCAGTGAAGAGCGATGTCAGCTAAT 8129
QY 62 ProIlePheProPheAspPheLysLeuGlySerGlyGluAspArgGlyLysValIle 81
Db 8130 CCAATATTCCATTGACITTAATTAGGTAGTTCAGGGAGGATAGAGGTAAGTTATA 8189
QY 82 ValThrGlnAsnGluAsnIleValTyAsnSerMetTyArgLysPheSerIleSerPhe 101
Db 8190 GTAACCCAGAAATGAATATTGTATATAATCTATGATGAAAGTTTGTAGCATTAGTTT 8249
QY 102 TPIleArgIleAsnLysTrpValSerAsnLeuProGlyTyThrIleIleAspSerVal 121
Db 8250 TCGATTAGNAPAAATAGGTAGTAAATTTACTCGATATCTATAATGTAGTAGTT 8309
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 8310 AAAATAACTCAGGTGCGATAGTATAGTATTATTAGTAAATTTTGTAGTATTCTTAAAA 8369
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyArgPheSerAsnAlaPro 161
Db 8370 CAATAGAGATAGTGACAAAGATATAATTTTAGTTATGATATATCAATATAGTCCT 8429
QY 162 GlyTyAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 8430 GGATACAATAAATGCTTTTGTAACTGTTACTAACAATATGATGGCAATATGAAGATT 8489
QY 182 TyrlleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 8490 TATATAATGGAAATTAATAGACTATAAAGTTTAAAGAACTAACTGGAATTAATTTT 8549
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 8550 AGCAAACTATAACATTTGAATTAATAAATAATTCACAGTACCGGTTGATTCTCAGAT 8609
QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrllePheAlaLysGluLeuAspGly 241
Db 8610 TCTGATAACATCAATCTGTGATAGAGATTTTATATATTTGCTAAAGAAATAGATGTT 8669
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValValLysAspTyTrp 261
Db 8670 AAGATATTATATATTATTAAGTTCGAATATCTAATGTTGTAAGATATTG 8729
QY 262 GlyAsnAspLeuArgTyAsnLysGluTyTrpMetValAsnIleAspTyTrpLeuAsnArg 281
Db 8730 GGAAATGATTTAAGATATAATAAAGAAATTTATATGTTTAAATAGATTATTAATAGA 8789
QY 282 TyrlleTyrlleAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301
Db 8790 TATATGATGCAACTCAGCAAAATGTTTATACACGTAGAAATATATAGACTTC 8849
QY 302 AsnGluGlyTyLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321
Db 8850 AATGAAGATATAAATAATTAATAAAGAAATCAGAGAAATCAAAATGATAGAGTA 8909
QY 322 ArgGlyGlyAspIleLeuTyrllePheAspMetThrIleAsnAsnLysAlaTyrlleAsnLeuPhe 341
Db 8910 CGAGAGAGATATTATTTATTTGATGACAAATTAATACAAAGCATATAATTTGTTT 8969
QY 342 MetLysAsnGluThrMetTyrlleAsnAspAsnHisSerThrGluAspIleTyrlleGly 361

```

8970 ATGAAGATGAACACTATGTCAGATAATCATAGTACTGAGATATATATGCTATAGGT 9029  
 QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
 Db 9030 TTAAGAGAACAAACAAAGGATATAAATGATAATATATATTTCAATACACCAATGAAT 9089  
 QY 382 AsnThrTyTrpTyrlleAsnSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
 Db 9090 AATACTTATTATTCGATCATCAAAATTTTAAATCAAAATTTTAAATGAGAAATATTTCT 9149  
 QY 402 GlyIleCysSerIleGlyThrTyArgPheArgLeuGlyGlyAspTrpTyArgHisAsn 421  
 Db 9150 GGAATATTTCAATAGGTACTTATCGTTTAGCTTGGAGTGATGGTATAGACACAAT 9209  
 QY 422 TyrlleValProThrValLysGlnGlyAsnTyrlleAsnSerLeuLeuGluSerThrSerThr 441  
 Db 9210 TATTGGTGCGCTACTGTGAACCAAGGAATTTAGCTTCATTTAGTAATCAACATCAACT 9269  
 QY 442 HisTrpGlyPheValProValSerGlu 450  
 Db 9270 CATGGGGTTTTGTACCTGTAAGTGA 9296  
 RESULT 6  
 AB061780 11747 bp DNA linear BCT 06-NOV-2001  
 LOCUS Clostridium botulinum orf-22, ha-70, ha-17, ha-33, ntaha, nt genes  
 DEFINITION for ORF-22, HA-70, HA-17, HA-33, NTNHA, neurotoxin, complete cds.  
 ACCESSION AB061780  
 VERSION AB061780.1 GI:16580753  
 KEYWORDS  
 SOURCE Clostridium botulinum (strain:C-Yoichi) DNA.  
 ORGANISM Clostridium botulinum  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.  
 REFERENCE 1  
 AUTHORS Sagane,Y., Kouguchi,H., Watanabe,T., Sunagawa,H., Inoue,K.,  
 Fujinaga,Y., Oguma,K. and Ohyama,T.  
 TITLE Role of C-terminal region of HA-33 component of botulinum toxin in  
 hemagglutination  
 JOURNAL Biochem. Biophys. Res. Commun. 288 (3), 650-657 (2001)  
 MEDLINE 21534265  
 REFERENCE 2 (bases 1 to 11747)  
 AUTHORS Sagane,Y., Watanabe,T., Kouguchi,H., Morita,M., Miyamoto,T.,  
 Yamamoto,T., Muroga,A. and Ohyama,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-MAY-2001) Toshihiro Watanabe, Tokyo University of  
 Agriculture, Faculty of Bioindustry, Yasaka 196, Abashiri, Hokkaido  
 099-2493, Japan (E-mail:t-watanabe@bioindustry.nodai.ac.jp,  
 Tel:81-152-48-3843, Fax:81-152-48-3843)  
 FEATURES  
 Location/Qualifiers  
 1..11747  
 /organism="Clostridium botulinum"  
 /strain="C-Yoichi"  
 /db\_xref="taxon:1491"  
 6..545  
 /gene="orf-22"  
 6..545  
 /gene="orf-22"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ORF-22"  
 /protein\_id="BA871744.1"  
 /db\_xref="GI:16580754"  
 /translation="MNDLFYAIENLKHNDNQNFNFIEMSLKKYIKETSKYNNLYDYNN  
 DILYHLWKELTEINLKNFSELDLRYISTKRYCINICKKKDKKIYINSEVYK  
 KLDVNVVSLYCNCFEFLDLISLTKYKQIIYMKFFEGRKDNEIAIRLSRQSIYK  
 IRIKSLKKLYPIVMQLVNI"  
 complement(620..2491)  
 /gene="ha-70"  
 complement(620..2491)  
 /gene="ha-70"  
 /codon\_start=1  
 /transl\_table=11





```

REFERENCE
AUTHORS      1 (bases 1 to 12297)
Hauser,D., Eklund,M.W., Boquet,P. and Popoff,M.R.
TITLE        Organization of the botulinum neurotoxin C1 gene and its associated
              non-toxic protein genes in Clostridium botulinum C 468
JOURNAL      Mol. Gen. Genet. 243 (6), 631-640 (1994)
MEDLINE      94301293
PUBMED      8028579
REFERENCE    2 (bases 1 to 12296)
Hauser,D.F.
TITLE        Direct Submission
JOURNAL      Submitted (03-MAY-1993) D.F. Hauser, Inst. Pasteur, Unite des
              Toxines Microbiennes, 28, Rue du Docteur Roux F-75724, Paris Cedex
              15, FRANCE
COMMENT      On Jul 27, 1994 this sequence version replaced gi:509268.
              Related sequences: X33751 & X66433.
FEATURES
source      1. .12297
              /organism="Clostridium botulinum phase IC"
              /strain="468"
              /sub_species="type C"
              /db_xref="taxon:35348"
              /lab_host="Clostridium botulinum"
              289..294
              /note="ORF22"
              302..841
              /note="ORF22"
              /codon_start=1
              /transl_table=11
              /protein_id="CAA51308.1"
              /db_xref="GI:509269"
              /db_xref="SPTREMBL:Q38195"
              /translation="MNDLFVAENLKHNDQHFNIEMSLKKYIEKTSKKYNLYVDYVN
              DILYHKKLEIINLKNSELDLSKYISTSIKVCINICKKKRDKKLIYNSEVYTK
              KLDVAVSYLYCONFEFLDILSLNYKEKQIYYMKFEGRKDNEIARLRLSRQSIYK
              IRTISUKKLYPIVMQLVNI"
gene         complement(854..2787)
              /gene="ANTP-70"
terminator   854..906
              /note="putative; ORF22"
              complement(854..906)
              /gene="ANTP-70"
terminator   /note="putative"
              /gene="ANTP-70"
              /note="putative"
              complement(916..2787)
              /gene="ANTP-70"
              /codon_start=1
              /transl_table=11
              /protein_id="CAA51309.1"
              /db_xref="GI:509270"
              /db_xref="SWISS-PROT:P46085"
              /translation="MSLSKELYITPKDS INNVNLADGNVYVNRGDGWLRSQNQLG
              FSTNSNGCAIVGDIRETATPYPTAFSENEYIKNNVQNFANFTEASEIPIGE
              FNTASPSKSLMYLQIYIRVEITKVLQNTVTERAVLYPSCLYKSIENFSEEGID
              KNYEFTSDQCKILNEKFTYKIDDTITVKESNNNNINENTSQTLIPYENGLYVINK
              GDGVMTDKDILGTLTLESSTSGSI IOPRLNTRTLPLENTSNPTIFSQVTEARLND
              AFNIQLPNTSTILFVEVEAPAKNKISMKYVNTYEKELINYQNGNIDKAEYLLPSL
              KCEPSTONNACVLPKVNTGVYKITTKNLPPLIIYEAGISSNRNNNSNNLSNDNIK
              AIKYITGLNRDAKLYLVSPFKNNYVIRIPISSSTISQLIFKRELGNISLADST
              VNILDLNLSGTHYTRQSPVDGNIYSQTLIPGDFNNIASSIFSFTRRNQOIGTLY
              RLFTESINGNLITINNYSDLLNNVEPISLLNGATVIFRKVKYTELNNYNIIFDAYRNS"
              complement(2801..3252)
gene         /gene="ANTP-17"
              complement(2801..3241)
              /gene="ANTP-17"
              /codon_start=1
              /transl_table=11
              /protein_id="CAA51310.1"
              /db_xref="GI:516172"
              /db_xref="SWISS-PROT:P46083"
              /translation="WSSRTFLPNNGYKIKSLFSDSLYLYVSSGALSFSNTSSLDNOK
              WKLEYLSSNGFRFNAPKPKLYANDYGFYILSSSSNNLSLWNPRIKAINSYIICITL
              SIYVNDYAWTVIYDNNNITDQIILNLPFNFINNSQILKLEKL"
              complement(3247..3252)
RBS

```

RESULT 7  
CBCBONT  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM



```

/ gene= "ANTP-17"
complement(3304. .4176)
/ gene= "ANTP-33"
complement(3304. .4164)
/ gene= "ANTP-33"
/ codon_start=1
/ transl_table=11
/ protein_id="CAA51311.1"
/ db_xref="GI:516173"
/ db_xref="SPTREML:Q38196"
/ translation="MSQTANDLRNNEVFVSPSNNNTKVLDKISQSEVKLNKLGGAG
NOKRLGIDYTKQKAYKIKMDNTSLITWNPASSVSVKTDITGNDQYLLQNYLSR
NVIIRNPNPLVQYIDIDTLDVYSTQSSNOFFKFSNCIYEALNRNCKLQTLNS
DRFSKLNLSQIIVLQWQFSSROKWIIEYNETKSAITLCKQENNYLFWIQNSNVV
ETVOSTSLIQWNIYLDNDASKIILYNQDINRVLDVYNSQIANGTHVIVDSYHGN
TNOQWLNLIY"
complement(4171. .4176)
/ gene= "ANTP-33"
4313. .8017
/ gene= "ANTP-139"
4313. .>8017
/ gene= "ANTP-139"
4411. .4416
/ gene= "ANTP-139"
4427. .8017
/ gene= "ANTP-139"
/ codon_start=1
/ transl_table=11
/ protein_id="CAA51312.1"
/ db_xref="GI:516174"
/ db_xref="SPTREML:Q38197"
/ translation="MDINDLNINSPVDNKNVIVRARKNTNTFFKFKVAPNITVAPE
RYGEPLDIAEYKLDGIDYDSNPLSDSERENFLQAIITLKRINNTISGKQLLSLI
STAIPFYGYGSSPNFTFKTKSNKLNLSVTSTIPPPFGYRETYIISQN
KNKYASVNIYFGGSNIVNNVYIYKKNDANGMTMAEIVFQPLLYKYNFYIDP
AMETLKLIKLYFLYGIKPSDNLVVPYRLTELDANKQESQLNIIDLLSGVDLEFI
NLTWPYTNSEFNPISIMFKYKNYKYTEEGNAGNDIKFLKQFQINQVQDIWL
NLNFCQSPNSIIPDRESNALKHFKQYTYMDYTDNYNINGVNFCINIKPLSNKN
TNISPKPVYNNVNRNISMKNYIDGDKGTTFDFTSYKIPNESEYERFNDSO
NPPNNISIEVDSPILDIINPYKNSDLVQTQTSMTTEVTHTALSINYLQAI
TNENFLSDFSKVSQKSLVYFLDNLMSYLETIKNDGPIDTDKRYILWLKEVF
KNYSFDNLTOEIDMCGINEVWFCALNLTNSFVEEYQDSGALSLSKQDL
REPNIIDIDSLLGLSFQDLNKLKLYISKNVYFKIYFSLFQWTEYYSOYFE
LICMAKQSLAESLVQIQVQNKFTDLSKASIPDPLKLIRETTEKTFIDLNSQIS
MYVDNLNKAICQVVEDLYPKFISYKXKYNINIKREFIQRCTININDEKSLI
NSYFKIDFKLDIQISIKNFFNSQVQVWKEILSQLELFLASKGPNILIEDSGK
NTLIQYTESIELYGVNGESLYLKSPPRETIKFSNFTGLTNFTICFWLRTGKND
DKTLIGNKVNCCWEIYFEDNGLVPEIIDSNGQESVYLSNIINDNWTYIISYDRL
KDQLIINDKNVANVSIDQILSYSTNIIISLVNKNNSIYVEELSVDNPITSEVIR
NFFSYLONSYTRDSKSLLEVKNQYLYVFPETSLYEVNDNKNKSYLSIKNTDGINI
SSVFKLINIDESXVYQKWDCTICVLDGTEKYLDISPENNRIQLVSSKDNKKITV
NTDLPRDCTIFSYNDXYFSLSLRDGDPNMIQNDNKNKPKGAHLILIS"
7931. .12032
/ gene= "BONT/C1"
7931. .>12032
/ gene= "BONT/C1"
8016. .8021
/ gene= "BONT/C1"
8032. .11907
/ gene= "BONT/C1"
/ codon_start=1
/ transl_table=11
/ protein_id="CAA51313.1"
/ db_xref="GI:516175"
/ db_xref="SWISS-PROT:P18640"
/ translation="MPTINNKYSDPDVKNKILYLDTHLNTIANEPKAFRTGNTW
VIPRFRSNPNLNPKNPTSPKSGYDNYENLSTSDSDTFLKLIKFLKRLNSREI
GEELYLSTDIPFGNNPTINTDFDVFNSVDVKTROGNWVXIGSINSLVITIG
PRENIDPETSTFKLTNTAAQGFQALSIIISPRFMTYSNATNDVGEGRFSKE
FCMDPILITMELHAMHNYLGTAPNDQTISSVTSNIFYSOYNVLEIYAFQSG
TIDILPKSARKYFEKALDYRISIAKRLNSITANPSSFNKYGEYKOKLIKRYFRVY
ESSEVTYNNKFLYELNLTQITFEFNKAIYNQNRKIYLSNVTYPYNTLDDNV
YDIQNGFNIKPSNLNLFMQNLNRNPAKRVAPENKMLYLFTEFKHAKIDGSRLYNK
LDCRELLAVKNTDLPFIFGDISDKVTDLFLRKDINEETEVIYYPDNVSDVDVILSKNTSE
HCQDLDLPSIDSESEIFLANKENAGVOGFLMWANDVVEDETTNLRDDEFTFTR
SIEEALDINSARVITYPTLANKNAGVOGFLMWANDVVEDETTNLRDDEFTFTR
VSAIIPGALNINSNRGNTEFAVGTGVTILLAEPEFTIPALGAFVYSKVOE
RNEIYKIDNCRLEQIRKRWKDSWMMGTWLSRIITQFNINSYQMTDSINYAGAIKA
KIDLEYKKGSGSKENIKSOVENKLSLDVKISEAMNNINKFIRECSVYLFKNMLPK
VIDELNEFDNRNTAKLINLIDSHNIILVGEVDKLKAKVNSFONTPFNIFSYTNLSL
LKDIIINEYFNINDSKILSLQNRKNTLVDTSYNAEVSEREGVDOLNPIPPDPKLGSS
GEDRGKVIYQENIYVNSYESFISFWIRINKVNSNLPGLYIIDSVNNSGWSIGI
ISNLFVTLQCNEDSEQISNFSISNNAAGYNKWFVTVNNMGNNKIIYINGKLID
TIKYKELTGNFSKITITFEINKIPDTGLITSDSNINMWIRDFYIFAKELDGKDIL
FNSQYTNVVKVDYNDLRKYKYVWYIDYLNRYANRSQIVFTIRNNNDNFEGY
KIIIKIRRGNTDTRVGGDILLYFVNDYLNKAYLNFKMNETMADHSEDYIAGLR
EOTDKINDNTIFQIQPMNNTYVVASOIFKNSNGENISGICSGTGYRFRGLSGDWRHN
YLVFTVQGNVASYLLESTSTHNGFVPSV"
11993. .12032
/ gene= "BONT/C1"
terminator
BASE COUNT 4825 a 1287 c 1703 g 4482 t
ORIGIN
Alignment Scores:
Pred. No.: 9, 23e-152 Length: 12297
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 7 Gaps: 0
US-09-910-186a-10 (1-450) x CBCBONT (1-12297)
QY 2 ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleAsn 21
Db 10558 ACAATACCCCTTAAATATTTTCATATACTAATAATCTTTTAAAGATATAATTAAT 10617
QY 22 GluTyPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 10618 GAATATTCATATATTAATTAATGATTCAAAAATTTTGACCTACAAACAGAAAAAFACT 10677
QY 42 LeuValasPThrSerGlyTyThrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 10678 TTAGTGGATACATCAGGATATAATGACAGAGTGATGAGTGAAGGCGATGTTCAGCTTAAT 10737
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 10738 CCAATATTTCCATTTGACTTTAAATTAGGTAGTACGGGAGGATAGAGTAAGTTATA 10797
QY 82 ValThrGlnAsnGluAsnIleValTyThrAsnSerMetTyThrGluSerPheSerIleSerPhe 101
Db 10798 GTAAACCCAGAAAGAAAAATTTGATATAATCTATGATGAAAGTTTTCAGCATTAGTTT 10857
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyThrIleIleAspSerVal 121
Db 10858 TGGATTAGATTAATAATGGTAAAGTAACTTACTGGATATATACTAATTAATGATAGTGT 10917
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 10918 AAAATAACTCAGGTGGAGTATAGGTATATTAGTAATTTTATTAGTATTACTTTAAAA 10977
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyThrAspIleSerAsnAsnAlaPro 161
Db 10978 CAAATGAAGATAGTAGCAACAGATATAAATTTTATGATATATCAATAATGATGCTCT 11037
QY 162 GlyTyThrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 11038 GGATAACAATAATGCTTTTGTAACTCTTACTAACAATAATGATGGGAAATATCAAGATT 11097
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 11098 TATAATAATGGAATAATTAATGATATCTATAAAGTTAAAGAACTAACTGAATTAATTTT 11157
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 11158 AGCAAACTATACCATTTGAATAATAATAATCCAGATACCGGTTTGTACTTACTTCAGAT 11217

```

Qy 222 SerAspAsnIleAsnMetTrrPleAArgAspPheTyrllePheAlaLysGluLeuAspGly 241  
 Db 11218 TCTGATACATCAATATGTCGATAAGAGATTTTATATATTTGCTTAAGAARTAGTAGGT 11277  
 Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGluTyThrAsnValValLysAspTyrrTp 261  
 Db 11278 AAGATATTAATATATTAATGATGTCGATATCAATATCAATGTTCTGAAGATATTGG 11397  
 Qy 262 GlyAsnAspLeuArgTyrrAsnLysGluTyrrMetValAsnIleAspTyrrLeuAsnArg 281  
 Db 11338 GGAATGATTAAAGATATAAAGAAATATATATGTTTAAATATGATTATTAAATAGA 11397  
 Qy 282 TyrMetTyrrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
 Db 11398 TATATGATGCGACTCAGCAAAATGTTTAAATACAGTAAATATATATGATGATC 11457  
 Qy 302 AsnGluGlyTyrrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
 Db 11458 AATGAAGGATATAAAATATATAAATAAAGAAATCAGAGGAATACAAATGATCTAGAGTA 11517  
 Qy 322 ArgGlyGlyAspIleLeuTyrrPheAspMetThrIleAsnAsnLysAlaTyrrAsnLeuPhe 341  
 Db 11518 CGAGGAGGAGATATTTATATTTGATATGACATTAATTAACAAGCATATATTTGTT 11577  
 Qy 342 MetLysAsnGluThrMetTyrrAlaAspAsnHisSerThrGluAspIleTyrrAlaIleGly 361  
 Db 11578 ATGAAGATGAACATATGTCAGATAATCATAGTACTGAAGATATATATGCTATAGT 11637  
 Qy 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
 Db 11638 TTAAGAGAACAAACAGGATTAATTAATTAATTAATTTTCAATATACCAACCAATGAAT 11697  
 Qy 382 AsnThrTyrrTyrrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
 Db 11698 AATACCTATATATAGCATCTCAAAATATTAATCAAAATTTTAATGAGAGAAATATTTCT 11757  
 Qy 402 GlyIleCysSerIleGlyThrTyrrArgPheArgLeuGlyAspTrpTyrrArgHisAsn 421  
 Db 11758 GGAATATGTTCAATAGGTAATCTATCTGTTAGACTTGAGGATGATGATAGACAAAT 11817  
 Qy 422 TyrLeuValProThrValLysGlnGlyAsnTyrrAlaSerLeuLeuGluSerThrSerThr 441  
 Db 11818 TATTTGGTGGCTACTGTGCAAGCAAGAAATATGCTTCATTATTAAGATCAACATCACT 11877  
 Qy 442 HisTrpGlyPheValProValSerGlu 450  
 Db 11878 CATTCGGGCTTTGTACCTGTAAGTGAA 11904  
 RESULT 8  
 20037920 11631 bp DNA linear BCT 24-JAN-2002  
 SCUS Clostridium botulinum genes for ORF-22, HA-70, HA-33, NTNEA,  
 DEFINITION neurotoxin, complete cds.  
 ACCESSION AB037920  
 VERSION AB037920.1 GI:6939789  
 KEYWORDS neurotoxin; NTNEA; HA-33; HA-17; HA-70; ORF-22.  
 SOURCE Clostridium botulinum (strain:D-4947) DNA.  
 ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.  
 1  
 Kouguchi,H., Watanabe,T., Sagane,Y., Sunagawa,H. and Ohyama,T.  
 AUTHORS In vitro reconstruction of the Clostridium botulinum type D  
 TITLE progenitor toxin  
 JOURNAL J. Biol. Chem. 277 (4), 2650-2656 (2002)  
 MEDLINE 21659747  
 2 (bases 1 to 11631)  
 Sagan,Y., Watanabe,T., Kouguchi,H., Yamamoto,T., Takizawa,J.,  
 Kabe,T., Murakami,F., Muroga,A., Nakatsuka,M. and Ohyama,T.  
 AUTHORS Direct Submission  
 TITLE Submitted (01-FEB-2000) Tohru Ohyama, Tokyo University of  
 JOURNAL Agriculture, Faculty of Bioindustry, Yatsuka 196, Abashiri, Hokkaido

099-2493, Japan (E-mail:t-oyama@bioindustry.nodai.ac.jp,  
 Tel:81-152-48-3838(ex.368). Fax:81-152-48-2940)  
 Location/Qualifiers  
 1. .11631  
 /organism="Clostridium botulinum"  
 /strain="D-4947"  
 /db\_xref="taxon:1491"  
 6. .545  
 /gene="orf-22"  
 6. .545  
 /genes="orf-22"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ORF-22"  
 /protein\_id="BAA90656.1"  
 /db\_xref="GI:6939790"  
 /translation="MNDLEVALENLKHNDHEDLEMSLKVKYIEKTSKYNLYDYVN  
 DLYHMKELIENKENSELDLRKYISTSKKVCINICKKKDKKLIYNSATYK  
 KLEAVNYSLYCEDFEFLDLISLNKKEQIIYMFECRKDNEIARHLHSRSIYK  
 IRIKSLKLYPIVMQLVNI"  
 complement(621..2492)  
 /gene="ha-70"  
 complement(521..2492)  
 /gene="ha-70"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HA-70"  
 /protein\_id="BAA90857.1"  
 /db\_xref="GI:6939791"  
 /translation="MSLSIKELYTKDKSINNVLADGNVYVNRGDGWLRSQNLG  
 GTSNNGCTAIVGDLRIRETATPYYPPTASENEEYIRNNVQNVFANFTEASEIGIGE  
 KFTASPNKGLYMYLOYTYRYEILKVLNTVIERAVLYVPSLGVAKSIFENSEQID  
 KNFYTSDEKILNEFIKKIAETTKAESDNNNTNLTATSTQILPYNGLYVINK  
 GDGYMTNDKDLIGTLIETISGSIQIPLRNTTRPLNTSNPLFSCFETEARLND  
 AFNIQENTSTLTFKEVEAPDNKSMKAYNTYKELINYNQNGNIADKAEYILPSL  
 GKEVSDAPSPQAVPVEVTEGFIQTPGNEIIVGVINPSENFEISTIPDDTY  
 NITONNACVLFVTNVTGVYKINAQNNPLIYIESIGSDNNNIQSNTLSNNIK  
 AINVTGDSNAESVLIYSLIKKNYIRIPIQISSSTNOLIFKRELGNISDLANST  
 VLIDNLTSGHYHTROSPDVGNKYSVLQITIPGDNNTASSIFSFRTRNQIGTLY  
 RLDESINGNLIIYKNSDILNNVPEISLLNGATYIFVKVYTELNNYIIFDAYNS"  
 complement(2506..2946)  
 /gene="ha-17"  
 complement(2506..2946)  
 /gene="ha-17"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HA-17"  
 /protein\_id="BAA90658.1"  
 /db\_xref="GI:6939792"  
 /translation="MSSERTFLPNGNYKISLFSLSLYTYSSGSLSTANTSSLDNQK  
 WKLEYISSNGPFRFSNVAPNPKYLAYNDYGYFLSSSSNNSLNWPDKIAINSYICTL  
 SIVNVDYAWTIYDNNNTIDQILNLPNFDINNSNQLKLEK"  
 complement(3008..3868)  
 /gene="ha-33"  
 complement(3008..3868)  
 /gene="ha-33"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HA-33"  
 /protein\_id="BAA90659.1"  
 /db\_xref="GI:6939793"  
 /translation="MSQTANDLRNNEVEFFISPSNNTNKVLDKISQSEVKLNKLSGA  
 NQWRLYTDNKQATYKVDNNTSLILWADPLSSVSVKTDNTNINQWYLQDYSR  
 NVILRNMPNLVQYNTDQTLIVSTQTNSSNQFFKSNCIYEALNNRCKLQTLNS  
 DRFLSNLNSQIIVLQWDFDSRQKWTIEYNKTSAYTLKQCNRYTLTQNSNNY  
 ETQYSDSLIQYWNINYLNDASKYILYNLQDTRVLVDVYNSQTANGTHVIVDSYHG  
 TNOQWIIINLI"  
 4129. 7719  
 /gene="ntnha"  
 4129. 7719  
 /gene="ntnha"  
 /codon\_start=1  
 /transl\_table=11

```

/product="NTNHA"
/protein_id="BAA90660.1"
/db_xref="GI:6939794"

/translation="MDINDDLINSPVDNKNVTVIRARKNTTFFKAFKVPAINWVAPE
LYRGSPLOIAERYKLDGQYDNLGDSERENFLOAIILLKRLNNTISGKQLLSLI
STAIPFPGYSSNPITFCKTKPSKNDKLSNLTVPFPFGYRETNISQN
KNFVASNIIVFGPGSNIVENNVCYKKNDAENGCMGTABILFQPLITYYKNYKFI
AMELAKCLIKSLFLYGLKSDDLVWPYRLTELDNKKPSQLMIDLLISGGVDLEFI
NTNPWFNYSFNSIKMFEKKNITYETEIGNNAIGNDKLKKOKFONKSVDLWNL
NUNYSKFNISIPDRFSAUKHETKQYTYTTEIGNNAIGNDKLKKOKFONKSVDLWNL
THIIKPEKVALNVAISLMKSNIVYDGLKGTEDFYSTKYPIYNEEYERYFNDS
PPLNNISIEEVDYSPEIIDNPYKXNSDOLLFTQITSTTEEVITHTALPVPNYLQAI
IKNEPNTSSDFSKVYSKSLSVYSPFLDNLMSYLEIKNDGSDTKKYVYLWAKVEI
TYNEPNTQIBDISGGINENYVFWFKALNIVLSTNSFVEEYQNSGSLISKKDNL
SPENATIDIDPSLLGKFSKDLNKLVELYISKNVPRKTYFNFLDQWTEYYSQVE
LICMAKOSITLAEQSVVKOLIQNKETDLSKASIPDLILKTEKTEFIDLNSEQIS
MRVDFNFKASICVVEDIIPKISISHEKIIINNINIKITREFORTCINIDNKSILLI
NLTITQIVSEVLVYGVNESIYLSKSPQVVMKESLPQLLLFATRGNISLEIDSGK
NDTRILGNKNCWEIYFDQGLVFIIDISNGQESVYLSNINNNWYIISISYDRL
KQDLIIFINDKNVANYSIQGLNIVYISIDSLNGKNSIYVEELSVDYRPAVESEIR
NYPFLYDINSYIDKSESLHKNVQLYNYVPETSLYEVDNKNKYSILKNTDGINI
PSVKFKLIDNDESKYGVQWDECIICVSDGTGYLDISPENNRIQVSSDKNAKKITV
NTDLPRDQCITSEYNDKTFESLSLDGQDYNMNCNDNNKVPKGAHLILAS"
/gene="nt"
/size="11591"
/CDs
/gene..11591
/size..11591
/codon_start=1
/transl_table=1
/product="neurotoxin"
/protein_id="BAA90661.1"
/db_xref="GI:6939795"

/translation="WTHPKYPNVSDPVNDNDLLLRIPQNKILITTPVKAFMITQNIW
VIPRPSSTNLSFKPRPKSKQSDPSYLSLTDQKDTFLKGLIKFLKRLNERNDI
GKGLINLVYSPFGMSDSTEDFTFTRHTNIAVEKFKNGSKWYNTIITSVLIFF
PCLNIDLYTASLTQGSQNSFFGFTLSILKVAPEFLTTFSDTSNSSAVLGKIS
FCMDPVIALHMLHLSHLQYGINIDPKRIPQVSGSEFTSQDQPNVQPEEYLTQFS
DVBIIPQIERLQRLREKALGHYKDIAKRLNINIKITPSMWSNIDKTKKIFSKYNPDK
DNTGNEVNDIFNLVSDLANVASEVYYSOYNKRWTFKSHYLPVPFANILDDNI
YITINGFNLTTKGLENLNGSINLRNPALOKLSSSVLDFTKVCLRLTRNSRDSSTC
IOVKNNLTPYVADKDSIQEITPESQITDENVENYSDNLSDELDAKATNPEN
DPLLPNNMPELNPVGESEYVDLITQVDVLNSYYLEAKOKLSNVENITLITTSVEE
ALGYSNKITVLPUSAEKYNQVAGFLFNANNEVEDFTTNIMKKDNLNVDLISDVSAI
IPYIGALNIGNSALRUKDQAFATVAGFLLEGFEFTIPALGVFTFYSDAQEREKI
ITENCLSRQYKRWKDSQYQWNSLSTRTFNHISYQMYDSLVSQADAKKIDL
EKYKYSQSGKENIKQVENLKNLSDYKISEAMNINKIFRECSVTYLFKNMLPKVVIDE
KNFLDKTUKTEILNIDSHNILLVEYDRLKAKVNESFENTIPFNFTVNNLSLKDMD
INEYFNINSDKILSQNKNUTMDTSYKNAEVEGNCVOLNIPFPDKIGSGGDDP
GKVTYQNSINVNMYKSEFSISFWIRIKKVNLCGYTLIDSKVKNSSQWSIGTISL
LUFYTLQKQNSQDINFSDISKKNAGYNKFFVITITNMGNMHIYINGKLIDITKY
KEUTGINLFTFOMKNIPNLTGSLDSSNINWIRKDTFPAKEDLNDNNGILFNLSL
QYTNVVKYDNGNDLIRKLYEINLVNMYNMYKSKNGIVFNTKKNNDNFNGEYKILI
KIRGTNDTRVGRGENLYFNFTINDKOYSLMGYKPSRNLGTDLPVLGALDQPMDEIR
KYGSTFIQCPNTFDYASQLSLSSNATNRLGILSIGSYFKLGDDYWFNHYLIPVI
KIEHYASLLESTSTHWVPVASE"
/size="4493"
/CDs
/gene="nt"
/size="4493"
/codon_start=1
/transl_table=1
/product="neurotoxin"
/protein_id="BAA90662.1"
/db_xref="GI:6939796"

/translation="WTHPKYPNVSDPVNDNDLLLRIPQNKILITTPVKAFMITQNIW
VIPRPSSTNLSFKPRPKSKQSDPSYLSLTDQKDTFLKGLIKFLKRLNERNDI
GKGLINLVYSPFGMSDSTEDFTFTRHTNIAVEKFKNGSKWYNTIITSVLIFF
PCLNIDLYTASLTQGSQNSFFGFTLSILKVAPEFLTTFSDTSNSSAVLGKIS
FCMDPVIALHMLHLSHLQYGINIDPKRIPQVSGSEFTSQDQPNVQPEEYLTQFS
DVBIIPQIERLQRLREKALGHYKDIAKRLNINIKITPSMWSNIDKTKKIFSKYNPDK
DNTGNEVNDIFNLVSDLANVASEVYYSOYNKRWTFKSHYLPVPFANILDDNI
YITINGFNLTTKGLENLNGSINLRNPALOKLSSSVLDFTKVCLRLTRNSRDSSTC
IOVKNNLTPYVADKDSIQEITPESQITDENVENYSDNLSDELDAKATNPEN
DPLLPNNMPELNPVGESEYVDLITQVDVLNSYYLEAKOKLSNVENITLITTSVEE
ALGYSNKITVLPUSAEKYNQVAGFLFNANNEVEDFTTNIMKKDNLNVDLISDVSAI
IPYIGALNIGNSALRUKDQAFATVAGFLLEGFEFTIPALGVFTFYSDAQEREKI
ITENCLSRQYKRWKDSQYQWNSLSTRTFNHISYQMYDSLVSQADAKKIDL
EKYKYSQSGKENIKQVENLKNLSDYKISEAMNINKIFRECSVTYLFKNMLPKVVIDE
KNFLDKTUKTEILNIDSHNILLVEYDRLKAKVNESFENTIPFNFTVNNLSLKDMD
INEYFNINSDKILSQNKNUTMDTSYKNAEVEGNCVOLNIPFPDKIGSGGDDP
GKVTYQNSINVNMYKSEFSISFWIRIKKVNLCGYTLIDSKVKNSSQWSIGTISL
LUFYTLQKQNSQDINFSDISKKNAGYNKFFVITITNMGNMHIYINGKLIDITKY
KEUTGINLFTFOMKNIPNLTGSLDSSNINWIRKDTFPAKEDLNDNNGILFNLSL
QYTNVVKYDNGNDLIRKLYEINLVNMYNMYKSKNGIVFNTKKNNDNFNGEYKILI
KIRGTNDTRVGRGENLYFNFTINDKOYSLMGYKPSRNLGTDLPVLGALDQPMDEIR
KYGSTFIQCPNTFDYASQLSLSSNATNRLGILSIGSYFKLGDDYWFNHYLIPVI
KIEHYASLLESTSTHWVPVASE"
/size="4493"
/CDs
/gene="nt"
/size="4493"
/codon_start=1
/transl_table=1
/product="neurotoxin"
/protein_id="BAA90663.1"
/db_xref="GI:6939797"

/translation="WTHPKYPNVSDPVNDNDLLLRIPQNKILITTPVKAFMITQNIW
VIPRPSSTNLSFKPRPKSKQSDPSYLSLTDQKDTFLKGLIKFLKRLNERNDI
GKGLINLVYSPFGMSDSTEDFTFTRHTNIAVEKFKNGSKWYNTIITSVLIFF
PCLNIDLYTASLTQGSQNSFFGFTLSILKVAPEFLTTFSDTSNSSAVLGKIS
FCMDPVIALHMLHLSHLQYGINIDPKRIPQVSGSEFTSQDQPNVQPEEYLTQFS
DVBIIPQIERLQRLREKALGHYKDIAKRLNINIKITPSMWSNIDKTKKIFSKYNPDK
DNTGNEVNDIFNLVSDLANVASEVYYSOYNKRWTFKSHYLPVPFANILDDNI
YITINGFNLTTKGLENLNGSINLRNPALOKLSSSVLDFTKVCLRLTRNSRDSSTC
IOVKNNLTPYVADKDSIQEITPESQITDENVENYSDNLSDELDAKATNPEN
DPLLPNNMPELNPVGESEYVDLITQVDVLNSYYLEAKOKLSNVENITLITTSVEE
ALGYSNKITVLPUSAEKYNQVAGFLFNANNEVEDFTTNIMKKDNLNVDLISDVSAI
IPYIGALNIGNSALRUKDQAFATVAGFLLEGFEFTIPALGVFTFYSDAQEREKI
ITENCLSRQYKRWKDSQYQWNSLSTRTFNHISYQMYDSLVSQADAKKIDL
EKYKYSQSGKENIKQVENLKNLSDYKISEAMNINKIFRECSVTYLFKNMLPKVVIDE
KNFLDKTUKTEILNIDSHNILLVEYDRLKAKVNESFENTIPFNFTVNNLSLKDMD
INEYFNINSDKILSQNKNUTMDTSYKNAEVEGNCVOLNIPFPDKIGSGGDDP
GKVTYQNSINVNMYKSEFSISFWIRIKKVNLCGYTLIDSKVKNSSQWSIGTISL
LUFYTLQKQNSQDINFSDISKKNAGYNKFFVITITNMGNMHIYINGKLIDITKY
KEUTGINLFTFOMKNIPNLTGSLDSSNINWIRKDTFPAKEDLNDNNGILFNLSL
QYTNVVKYDNGNDLIRKLYEINLVNMYNMYKSKNGIVFNTKKNNDNFNGEYKILI
KIRGTNDTRVGRGENLYFNFTINDKOYSLMGYKPSRNLGTDLPVLGALDQPMDEIR
KYGSTFIQCPNTFDYASQLSLSSNATNRLGILSIGSYFKLGDDYWFNHYLIPVI
KIEHYASLLESTSTHWVPVASE"
/size="4493"
/CDs
/gene="nt"
/size="4493"
/codon_start=1
/transl_table=1
/product="neurotoxin"
/protein_id="BAA90664.1"
/db_xref="GI:6939798"

/translation="WTHPKYPNVSDPVNDNDLLLRIPQNKILITTPVKAFMITQNIW
VIPRPSSTNLSFKPRPKSKQSDPSYLSLTDQKDTFLKGLIKFLKRLNERNDI
GKGLINLVYSPFGMSDSTEDFTFTRHTNIAVEKFKNGSKWYNTIITSVLIFF
PCLNIDLYTASLTQGSQNSFFGFTLSILKVAPEFLTTFSDTSNSSAVLGKIS
FCMDPVIALHMLHLSHLQYGINIDPKRIPQVSGSEFTSQDQPNVQPEEYLTQFS
DVBIIPQIERLQRLREKALGHYKDIAKRLNINIKITPSMWSNIDKTKKIFSKYNPDK
DNTGNEVNDIFNLVSDLANVASEVYYSOYNKRWTFKSHYLPVPFANILDDNI
YITINGFNLTTKGLENLNGSINLRNPALOKLSSSVLDFTKVCLRLTRNSRDSSTC
IOVKNNLTPYVADKDSIQEITPESQITDENVENYSDNLSDELDAKATNPEN
DPLLPNNMPELNPVGESEYVDLITQVDVLNSYYLEAKOKLSNVENITLITTSVEE
ALGYSNKITVLPUSAEKYNQVAGFLFNANNEVEDFTTNIMKKDNLNVDLISDVSAI
IPYIGALNIGNSALRUKDQAFATVAGFLLEGFEFTIPALGVFTFYSDAQEREKI
ITENCLSRQYKRWKDSQYQWNSLSTRTFNHISYQMYDSLVSQADAKKIDL
EKYKYSQSGKENIKQVENLKNLSDYKISEAMNINKIFRECSVTYLFKNMLPKVVIDE
KNFLDKTUKTEILNIDSHNILLVEYDRLKAKVNESFENTIPFNFTVNNLSLKDMD
INEYFNINSDKILSQNKNUTMDTSYKNAEVEGNCVOLNIPFPDKIGSGGDDP
GKVTYQNSINVNMYKSEFSISFWIRIKKVNLCGYTLIDSKVKNSSQWSIGTISL
LUFYTLQKQNSQDINFSDISKKNAGYNKFFVITITNMGNMHIYINGKLIDITKY
KEUTGINLFTFOMKNIPNLTGSLDSSNINWIRKDTFPAKEDLNDNNGILFNLSL
QYTNVVKYDNGNDLIRKLYEINLVNMYNMYKSKNGIVFNTKKNNDNFNGEYKILI
KIRGTNDTRVGRGENLYFNFTINDKOYSLMGYKPSRNLGTDLPVLGALDQPMDEIR
KYGSTFIQCPNTFDYASQLSLSSNATNRLGILSIGSYFKLGDDYWFNHYLIPVI
KIEHYASLLESTSTHWVPVASE"
/size="4493"
/CDs
/gene="nt"
/size="4493"
/codon_start=1
/transl_table=1
/product="neurotoxin"
/protein_id="BAA90665.1"
/db_xref="GI:6939799"

/translation="WTHPKYPNVSDPVNDNDLLLRIPQNKILITTPVKAFMITQNIW
VIPRPSSTNLSFKPRPKSKQSDPSYLSLTDQKDTFLKGLIKFLKRLNERNDI
GKGLINLVYSPFGMSDSTEDFTFTRHTNIAVEKFKNGSKWYNTIITSVLIFF
PCLNIDLYTASLTQGSQNSFFGFTLSILKVAPEFLT
```







```

NVLIRNWNPNVLQYNWDTLLVSTQSSNNOFFFSCNICEIYALNNRNCKLQIOLNS
DRFLSKNLQSIIVLWOWFDSQKWLIEYNETKSAITLKCQENNRNRYLTWIONSNNY
EYQSDLSLQYNNIYLDNDASKIYLLNLDQTNRLVDVINSQIANGTNVIVDSYHGN
TNOQIINLI"
4145..7735
/gene="ntaha"
4145..7735
/gene="ntaha"
/codon_start=1
/transl_table=11
/product="NTNHA"
/protein_id="BA089712.1"
/db_xref="GI:6729212"
/transl_table="MDINDOLNINSPVDNKNVIVPARKTNTFFKFKVAPNINWIAPE
RYGEPDIAEEYKLDGIVDSNPLSODSERENFLOAIITLLKRNINTISGKQLLSLI
STALPIYGVYGGSPNITFEKTPKKNKLSNLTSTIPFPFGYRETHYIESPN
NKDYAFNIVFGPNSVNNVYICKKNDANGMTMAEILFQPLLKYKNKFIIDP
AMELTKLISLXLYGKPSGLVYPIRELDNKPQSLNIDLLISGGVDEFI
NTPYFTNSFYNSIKMFYKNIYETELEGNAJNDIKRLRCKFONSQVDIWL
NLNYSKEFNSIPDSNALKHFKYKQYTMDDYDNINGVNGOINIKPLSDKN
TNISKEKVVNLVNNNISMKSNIYDGLKGTEDFYSTYKIPNIEYEFNDS
NPNLNSIEVDSIPDIIDINIKONSNLVTQITSMTEEVTHIALPILNLOAI
TNENFNLSDFSQSSGKSLVYFGLNMSIYLETIKNDGPDIDDKKYLWLKEVF
KNSYFNLDQIESGGINVYFGLNMIINTNSFEVYQNSGPTSLISKNDL
SEPIEDIDPDLGLSDFDLANKLYEYKNIYFKIYFNFLDQWTEYISQYFE
LICAKQSLIAQESLVCIOIKQFTDLKASIPDYLKIKETKFTDLKESQIS
NRYVDNLNASCIVFEDYIPKFIYSVMEKYINNINIKRERTORTNNDKESILI
NSYFKPIDKFDLQIKKFNFSQVOMKMLSPKOLLLEFATRGPSNLLIEDISGK
NTLQYVESLIVYNGESLYLKSNEIVEFNSFNTNGLNINFTICFLWAFKDD
DKTLIKGNKNWGEYIFEDNGLVFEIIDSNGNSVLSVINNNWYIYISVDRS
KQDLIFINKNVNAIEQILATYISNVLNKNNSYVSELSVDKTYISERVIR
NYFGLNDSYIRDSKSLLEYKNYQLYNFFPKTSUYEVNNDKNSLSKNDGINI
PSVFKLINDESGYVQKWDCEICVSDGTEKYLIDISENNRNIOLVSKDKAKTIV
NTDLFRPDCITFSYNDXVFSLSLDGSDYNNMCDNKNYKPKGAHLILES"
7750..11592
/gene="nt"
7750..11592
/gene="nt"
/codon_start=1
/transl_table=11
/product="neurotoxin"
/protein_id="BA089713.1"
/db_xref="GI:6729213"
/transl_table="MPTIINFNYSVDVDMKNLILYDHTNLNANPEKAFRIIGNIW
VIPRFRSDNPNLKNPRTYSPKSGPYDNYLSTSEKDTFLKEIKLFRINSREI
GEELIYRLATDIPPPGNNTPIINTFDVDFNSVDKTRQGNWVATGINSFVITG
PREMIDPSTPKLTNTNFAAOFGALSIIISPREMLTYSNATNNVNGEYRFSKSE
FCMDPILILHMLNHNLYGTAIPNDQRISSVTNIFYSQYKVKLEAYIYRGGP
TIDLIPSAKRYPEKALDYRSIAKRLNITITANPSSNKYIGEYKQILIKYRIV
ESSEVADRKNFELIKETIQTFFENAKIYNQNKIYLSNVIPTIANLDDNV
YDIONFNPISNLNVLFMQNSLRNPNALRKVNPERMLYLFKFCRKAIDGRSLYNT
LDCELYKNTDLPIDIGIDIDIDFSDKINEETEVYIDYDNDVQVQVILSKNISE
HGQDLIYPIIGESQVLPQENQVYDNRQNDVYLSNYYLESKLSDNVEDFTFT
SIEALDNGSVYTYFPEKADKVNQVGGGLFWANDVVEFTNILKRDLDKISD
VSATIPYPALNINSVRRCNTPFAVGTWILLEAPOEFTIPALGAFVITYKVOE
RNEILKTDNLEQRIKWDKSEVMIGTWSLRTTQFNINISYQMDVSLNQAQALD
KIDLEYKISSGRENKISQVENIKNSLDIKISEAMNNINKFRESYVILKMLPK
VIDELNKFDKTLKLNLDNSHILVGEVDRLKAKNFESFENTIPFNITNSNL
LKDINEFNSINDSKLSQKNKNAIVDSGYNAEVRLEDGVQVQVITNDFKSSS
GDKIIVLNNILNLYSAIVNSVSFWIKISDITNSHNEYTIINSKONGSKLCIRN
GNIWILQDINRKYSLIFDYSELSHTGTNKKWFFVTITNNIMYKMLYINGELKQS
ERIDNLEVKDKTIVGIDENIDENOMLWIRDFNIFSKELSNEDNIVYEGQILNV
KIDWGRPLKEDTEYIINDNKYIDRYTAPKSNILVLVQVDPDRSKLVTGPIKSVND
KNPYRSLKNGNIMFHLNYSKGKYMILRDFTIYIAREGRCRNCVYALKLQSNLGN
GIGIFISNIVQNKYCSQIFSPFMKNLMLLDIYKPRFSEFNATYPVAVTNIEKL
LSTSFWKIFSRDFGWVE"
BASE COUNT 4499 a 1240 c 1651 g 4224 t
ORIGIN

```

Alignment Scores:  
 Pred. No.: 2,618-50 Length: 11614  
 Score: 882.50 Matches: 198  
 Percent Similarity: 61.20% Conservative: 78

```

Best Local Similarity: 43.90% Mismatches: 148
Query Match: 36.91% Indels: 27
DB: 1 Gaps: 10
US-09-910-186a-10 (1-450) x AB037166 (1-11614)
QY 2 ThrileProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsn 21
DB 10276 ACAATACCCCTTTAATATTTTTCATATACATAAATCTTTTATTAAGAATAAATAAT 10335
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAspThr 41
DB 10336 GAATATTTCAATAGTATTAAATGATTCAAAATTTTGAGCTTACAAAACAAAAAATGCT 10395
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
DB 10396 TTAGTGGATACATCAGGATATAAATGCAAGAGTAGGCTAGAGGTGATGTCAAGTAAAT 10455
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
DB 10456 ACGATATATACAAATGATTTTAAATTAAGTAGTTCAGAGAT-----AAAATTATA 10506
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
DB 10507 GTAAATTTAAATAATAATTTTATATAGCGCTATTTATGAGAACTCTAGTGTAGTTT 10566
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyrThrIleIleAspSer 120
DB 10567 TGCATTAGATATCTAAGATTTAACTAATTTCTCATATGAATACATAATAATTAATAGT 10626
QY 121 ValLysAsnAsnSerGlyTyrTrpSerIleGlyIleLeuSerAsnPheLeuValPheThrLeu 140
DB 10627 ATAAACAAAATTTCTGGTGGAAATTTATGATTAGGAATGCAATATAGATGGATTTTA 10686
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
DB 10687 CAAGATATTAATAGAAAATATAAAGTTTAAATTTTGTATATAGTGAATCATTAAGTCAT 10746
QY 161 ProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet 179
DB 10747 ACAGGATATACAAATAAATGTTTGTCTACTATACTAATAATAATAATGCGGTATATG 10806
QY 180 LysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyLe 199
DB 10807 AAACCTTTATATAATGGAATTTAAAGCAGAGTGAAGAAATTTGAAGATTTAAAGAGCTT 10866
QY 200 AsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr 219
DB 10867 AAGTTAGATAAAACCATAGTATTGGATAGATGAGAAATATAGAT----- 10911
QY 220 SerAspSerAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeu 239
DB 10912 -----GAGAATCAGATGCTTTGGATTAGAGATTTTAAATTTTCTAAAGAATA 10962
QY 240 AspGlyLysAspIleAsnIleLeuLeuSerLeuGlnTyrThrAsnValValLysAsp 259
DB 10963 AGCAATGAGATATTAAATATGTTATATGAGGACAAATATTAAAGAAATGTTATTAAAGAT 11022
QY 260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279
DB 11023 TATTGGGAAATCCTTTGAAAGTTTGAATAGAGATTTTAAATTTTCTAAAGAATA 11082
QY 280 AsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn 299
DB 11083 GATAGTATATAGACCTTAAAGATAAATACTTGTCTGTTGTCAGTATCCAGATAGATCT 11142
QY 300 AspPheAsnGluGlyTyrLysIleIleLeuLysArgIleArgGlyAspThrAsnAspThr 319
DB 11143 AAATTATATCTGGAAATCTTATTACTATTAAATCAGTATCTGATGAAGAAATCCTTATAGT 11202
QY 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339
DB 11203 AGAATTTTAAATGGAGATAAATAATGTTTTCATATGTTATATATAGTGGGAAATATATG 11262

```



QY 340 LeuphemetylSAsnGluThrMetyrAla-----AspAsnHisserThrGluasp 356  
 DB 11263 ATAAAGAGATAGTATACATAATATGCAATAGAGAGAGAGTGTCAAAAAATGCT 11322  
 QY 357 IletYrAlaIleGlyLeuArgGluThrLysAspIleAsnAspAsnIleIlePheGln 376  
 DB 11323 GATATGATCAATAAATACAGAGTAATTTAGGTATAGT---ATATTTAGT 11379  
 QY 377 IlegInPrometAsnAsnThrTyTyTyAlaSerGlnIlePheTylsSerAsnPhAsn 396  
 DB 11380 ATAAAAATATGATGATCAAAATAATATGATGATCAATTTCTCTAGTTTATG--- 11436  
 QY 397 GlyGluAsnIleSerGlyIleCySerIle---GlyThrTyArgPheArgLeuGlyGly 415  
 DB 11437 ---AAAAATCAATGCTTCTAGCAGATATATAAACTTGGAGATTCTTTT--- 11487  
 416 AspTptTyArgHisAsnTyLeuValProThrValLysGlnGlyAsnTyAlaSerLeu 435  
 11488 -----GAAATGCATACACGCCA---GTTCCAGTAACATATATGACACAAA 11532  
 436 LeuGluSerThrSerThrHisTrpGlyPheVal 446  
 11533 CTATTATCAACTTCATCTTTTGGAAATTATT 11565  
 RESULT 12  
 CBNUTD 3952 bp DNA linear BCT 17-MAY-1995  
 DEFINITION Clostridium botulinum gene for neurotoxin type D.  
 ACCESSION X54254  
 VERSION X54254.1 GI:40395  
 KEYWORDS neurotoxin; neurotoxin type D.  
 SOURCE Clostridium botulinum.  
 ORGANISM Clostridium botulinum  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium  
 1 (bases 1 to 3952)  
 Niemann,H.  
 DIRECT SUBMISSION  
 TITLE Submitted (03-AUG-1990) Niemann H., Institut f Mikrobiologie,  
 Bundesforschungsanstalt f Viruskrankheiten, Postfach 11 49, D 7400  
 Tuebingen, F R G  
 2 (bases 1 to 3952)  
 Binz,T., Kurazono,H., Popoff,M.R., Eklund,M.W., Sakaguchi,G.,  
 Kozaki,S., Kriedelstein,K., Henschen,A., Gull,D.M. and Niemann,H.  
 Nucleotide sequence of the gene encoding Clostridium botulinum  
 neurotoxin type D  
 Nucleic Acids Res. 18 (18), 5556 (1990)  
 JOURNAL  
 MEDLINE 91016853  
 PUBMED 2216736  
 3 (bases 1 to 3952)  
 Okimoto,R., Macfarlane,J.L. and Wolstenholme,D.R.  
 Evidence for the frequent use of TTC as the translation Initiation  
 codon of mitochondrial protein genes in the nematodes, Ascaris suum  
 and Caenorhabditis elegans  
 Nucleic Acids Res. 18 (20), 6113-6118 (1990)  
 JOURNAL  
 MEDLINE 91045077  
 PUBMED 2235493  
 FEATURES  
 Location/Qualifiers  
 1..3952  
 /organism="Clostridium botulinum"  
 /strain="BVD/-3"  
 /db\_xref="taxon:1491"  
 31..35  
 /note="ribosome binding site"  
 47..3877  
 /note="neurotoxin type D"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA38175.1"  
 /db\_xref="GI:40395"  
 /db\_xref="SWISS-PROT:P19321"  
 /translation="MTWPKDFNYSDFVNDNLYLRIFQNKLIITPVKAFMITQNIW

VIPERSSTNPSLSKPPRTSKYQSYDPSILSTDEQKDFLKGILKFLKRIINERDI  
 GKXILNYLVGSPFMGDSSTPDTFTRHTTNIIVAKFPGSKVNTIITPVLIFG  
 PNLNIDYASLTQQQSNPSEFEGTSLIKVAPEFILFSDVTSNOSAVLGKSI  
 FCMPIALMHETLSHLQYGINPSOKRIRPQVSEGGFSDQGNVQFELYTFGGL  
 DVPIIPIERSOLREKALGHYDKIAKRNINKTIPSSGFWISNDIKYKIFSEKYNFKD  
 DMTGEVNLIDAFNSYLSLTVMSEVYSSOYNVKNRTHYFVRHLYLVEFVAILDDNI  
 YTRDGPNTKNGKTIENSQNIENPALOKTSSSEVVDLFTKVCRLTKNSRDDSTC  
 IKYNNRLPYVADKDSISEIFENKLIIDETNQNYSKESLDESILDGQVPIINPEIV  
 DPLPNVMEPLNGPEEIVFYDDITKYVDLYNLSYIYESQKLSNNENILITTSVEE  
 ALGYSNKIYTFPLSLAEKYNQVQAGFLNANVEVDFETNIMKDKDLDKISDVSVI  
 IPIYGALNIGNSALRGNFNOAFATAGVAFLEGGPEFTIPALGVTFYFYSIQEREKI  
 IKYIENCLQRVKRMDKQYMWNSLRTTFQFNHINYQMYDSLSYQADAIKAKIDL  
 EYKYSKDKENIKSOVENKSLDKVISEANNINKFRECSTVILEKMLPKVIDE  
 LNKFLRTTEILNIDSENILLVGEVDRLKAKVESFENTMPPNFYSYTNLSLLKDI  
 INEYFNSINDSKILSONKALVTSQYNAEVRVGDVQVNTIYTNDFKLSGSDKI  
 WYLNNNILYSAIENSYSVFWIKSKDNTNSHNEYTINSIQNSQKHLKIRNGNIE  
 IYNNNNILYKSLIFDYSSLSHTGTYTNKFFVTIINLMGYMKLYIINGELKQSKIE  
 IVDLDEKLTIVGIDENIDENQMLWIRDFNFSKELSDENINIVYEQIILRNKDY  
 WGNPLKFDTEYTIINDNYIDRYAPESNVVLVQVDRSKLYTGNPTIKSVSDKNPY  
 SRILNDNTILHMLYNSRKYMIIRDITDIYATQGGECQNCVYALKLQSLNGYIGI  
 FSKNIYKSKYCSQIFSSFERENTMLADIYKWPWFSEFKNATYPVAVTNYETKLLSTS  
 SFWKFTISRDPGWVE"  
 BASE COUNT 1597 a 398 c 610 g 1347 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9,24e-50 Length: 3952  
 Score: 867.50 Matches: 195  
 Percent Similarity: 61.64% Conservative: 83  
 Best Local Similarity: 43.24% Mismatches: 146  
 Query Match: 36.28% Indels: 27  
 DB: 1 Gaps: 11  
 US-09-910-186A-10 (1-450) x CBNUTD (1-3952)  
 QY 2 ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
 DB 2561 ACAATGCTTTTAAATATTTTTCATATACATAATATCTTTATTAAGATATAATAT 2620  
 QY 22 GluTyPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
 DB 2621 GAATATTTCAATAGTATTAATGATTCAAAAATTTTGAGCTTACAAAACAAAATGCT 2680  
 QY 42 LeuValAspThrSerGlyTyThrAsnAlaGluValSerGluGluGlyValGlnLeuAsn 61  
 DB 2681 TTAGGGATACATCAGATATAATGCAGAGTGGGGTAGGAGATATGTTCAACTTAT 2740  
 QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
 DB 2741 ACGATATATACAAATGACTTTAAATTAAGTAGTTTCAGGAGAT-----AAAATTATA 2791  
 QY 82 ValThrGlnAsnGluAsnIleValTyThrAsnSerMetTyThrGluSerPheSerIleSerPhe 101  
 DB 2792 GTAAATTTAAATAATAATATTTTATAGCGCTATTTATGAGAACTCTAGTGTATGTTT 2851  
 QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyThrIleIleAspSer 120  
 DB 2852 TGGATTAAAGATATCTAAAGATTTAACTAATCTCATAATGAATATACATAATAACAGT 2911  
 QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140  
 DB 2912 ATAGAACAAATCTGGGTGAAATATGATTAGGAATGCCAATATAGATAGGATTTTA 2971  
 QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyThrAspIleSerAsnAla 160  
 DB 2972 CAAGATGTTATAGAAAGTATAAAGTTTAAATTTTTCATTATAGTGAATCATTAAGTCAT 3031  
 QY 161 ProGlyTyThr---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet 179  
 DB 3032 ACAGCATATACAAATAAATGTTTGTACTACTAATAATAATAATGCGGTATATG 3091  
 QY 180 LysIleTyThrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199  
 DB 1991 LysIleTyThrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199



Db 3092 AACTTTATATAAATGGAATTAAGCAGAGTCACAAAATTTGAAGATTAGATGAGGTT 3151  
 QY 200 AspPheSerLysThrIleThrPheGluLeuAsnLysIleProAspThrGlyLeuIleThr 219  
 Db 3152 AAGTTAGATAAAACCATAGATTGGAATAGATGAGATATAGAT----- 3196  
 QY 220 SerAspSerAsnIleAsnMetIrpIleArgaspPheTyrIlePheAlaLysGluLeu 239  
 Db 3197 -----GAGAAACAGATGCTTGGATTAGAGATTTTAAATATTTTCTTAAGAAATTA 3247  
 QY 240 AspGlyLeuAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAsp 259  
 Db 3248 AGTAATGAAGATTAATATGATATGATGAGGACAAATATTAAAGAAATGTTATTAAAGAT 3307  
 QY 260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279  
 Db 3308 TATGGGGAAATCCCTTGAAGTTTCATACAGAAATATTAATTAATGAAATTAATATA 3367  
 QY 280 AsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsn 299  
 Db 3368 GATAGTATATAGCACCTGAAAGTAATGACTGTGACTGTGTTTCAGTATCCAGATAGATCT 3427  
 QY 300 AspPheAsnGluGlyTyrLysIleIleLysArgIleLysArgGlyAsnThrAsnAspThr 319  
 Db 3428 AAATTATATACCTGGAATCTTACTATTAAATCAGTATCTGATAAGAAATCCCTTATAGT 3487  
 QY 320 ArgValArgGlyLysAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339  
 Db 3488 AGAATTTTAAATGGAGATAATAATATCTTCATATGTTATATATAGTAGGAATATATG 3547  
 QY 340 LeuPheMetLysAsnGluThrMetTyrAlaAspAsnHis-----SerThrGluAsp 356  
 Db 3548 AATAAAGAGATAGTACTACATAATATGACACACAGGAGGAGAGTGTTCCAAAAATGT 3607  
 QY 357 IleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln 376  
 Db 3608 GTATAGCAATTAATACAGAGTAATTTAGTAATTTATGTTATAGT---ATATTAGT 3664  
 QY 377 IleGlnProMetAsnAsnThrTyrTyrTyrIleValSerGlnIlePheLysSerAsnPheAsn 396  
 Db 3665 ATAAAAAATATTGTAATAAATAAATATTGTAATAAATTTTC---TCTAGTTTATAGG 3721  
 QY 397 GlyGluAsnIleSerGlyLysSerIle---GlyThrTyrArgPheArgLeuGlyGly 415  
 Db 3722 ---GAAATACAAATGCTTACAGATATATATAAACCCTGGAGATTTC--- 3769  
 QY 416 AspTrpTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu 435  
 Db 3770 -----TTTAAAAATGCATACACGCGAGTGCAGTAACCT-----AATTATGAACACAAA 3817  
 QY 436 LeuGluSerThrSerThrHisTrpGlyPheVal 446  
 Db 3818 CTATTATCAACTCATCTTTTGGAAATTTATT 3850  
 RESULT 13  
 S49407  
 LOCUS  
 DEFINITION  
 type D neurotoxin [bacteriophage d-16 phi, host = C. botulinum,  
 type D, CB16, genomic, 4087 nt].  
 S49407  
 ACCESSION  
 VERSION  
 S49407.1 GI:260238  
 KEYWORDS  
 Bacteriophage d-16 phi CB16 host = C. botulinum.  
 SOURCE  
 Bacteriophage d-16 phi  
 ORGANISM  
 Viruses.  
 1 (bases 1 to 4087)  
 Sunagawa,H., Ohnaka,T., Watanabe,T. and Inoue,K.  
 The complete amino acid sequence of the Clostridium botulinum type  
 D neurotoxin, deduced by nucleotide sequence analysis of the  
 encoding phase d-16 phi genome  
 J. Vet. Med. Sci. 54 (5), 905-913 (1992)  
 JOURNAL  
 93042276  
 MEDLINE  
 PUBMED 1420572

REMARK GenBank staff at the National Library of Medicine created this  
 entry [NCBI gisseq 118639] from the original journal article.  
 This sequence comes from Fig. 4.  
 FEATURES  
 Location/Qualifiers  
 1. 4087  
 /organism="Bacteriophage d-16 phi"  
 /db\_xref="taxon:12409"  
 /note="type: D"  
 182..4012  
 /gene="type D neurotoxin"  
 182..4012  
 /gene="type D neurotoxin"  
 /note="Method: conceptual translation with partial peptide  
 sequencing; This sequence comes from Fig. 4"  
 /codon\_start=1  
 /product="type D neurotoxin"  
 /protein\_id="AAB24244.1"  
 /db\_xref="GI:260239"  
 /translation="MTVPKDFNYSDPVNDNDILYRIPQNKLTTPVKAFTITQNIW  
 VIPERFSSDTNPSLSKPRPTSKYQSYDPSYSLTDEQKDTFLKGIKLFKRNEDI  
 GKILNLYVSGPMGDSSTPDTDFRTTNIKVEKFGKMGKVTNITTSVLFIFG  
 PLNLYDTASLILQCOGSPGEGCTLSILKVAPEFLTFSDVTSSNOSSVLGKSI  
 FCDPVALMHEILHSLHLYGNIIPSDKRIKIRFQVSEGGFSDQGNVQREELTGGL  
 DVEIIPQERSQLREKALGHYKDKIANNINKITPSSWISNIDKAKIFSEKINFDK  
 DNTGNFVNIDKNSLYSDLTNVMSEVYVSYQYNVNRTHYFRHLPVFNILDDNI  
 YTRDGNLTNKGFIENSGQNIERNPALQKLSSEVDLTFTKVLRLTKNRDSTC  
 IKVNNRLPYVADKDSISQIFENKIITDNTVQNTSDNFSLESLDQGVLPETV  
 DPLPNVMEPLNLPGEIVFDYDITKYVDYLSYVYLSQKLSNNVENTITLTSVEE  
 ALGYSNKIYFPLSLAKVNGKYOAGFLANAVEVEDFTNLMKKDLTKLSDYSVI  
 IPIYGPALNIGNSALRNFQKATAGVAFELGEGPEFTIPALCVETFFYSIOERKI  
 IKTIENLEQKRWKDSIOVMVSNULSRITQFNHINTQMDSYQADAKAKIDL  
 EYKYSKDKENIKSQVENLKNLSDVKISEAMNNINKFRECSVTYLFKNMLPKVIDE  
 LNKYDLTKTELINLIDSHNIILVGEVDRLKAKVNESFENTMPTNFTNNLSLKI  
 INEYFNSNDISKILQNKRLNVDYTCYNAEVRVGDVQNTIYNDFKLSGSGDKI  
 IVNNNNLYSAIYENSVSFVKISKDLTNSHNEYTIINSIENSGKLCIRGNIE  
 WILDQVARKYKSLIEDYSELSHSGYTKWFEVTITNNGYMKLVINGELKOSQKIE  
 DLDEVKDLIIVGIDENIDENQMLWIRDNFISKELSNEDINIVYEGQILRVKDY  
 WGNPLDETEYIIINDYIAPESNVLVFRDPRDKLYTGNTITTSKVSQKNPY  
 SRILGNINILHMLYNSKRYMIRDTDTIYATGEGCSQNCVYALKLQSNLNGYIGI  
 FSIKNGYKKNKYCSQIFSFRENTMLADIYKPRFSEKNAVTPVAVTNYETKLST  
 SFWAFISRDPGWVE"  
 BASE COUNT 1646 a 413 c 633 g 1395 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9,56e-50 Length: 4087  
 Score: 567.50 Matches: 195  
 Percent Similarity: 61.64% Conservative: 83  
 Best Local Similarity: 43.24% Mismatches: 146  
 Query Match: 36.28% Indels: 27  
 DB: 7 Gaps: 11  
 US-09-910-186A-10 (1-450) x S49407 (1-4087)  
 QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsn 21  
 Db 2696 ACAATCCCTTTTAAATATTTTTCATATACTAATAATCTTTATTAAGAGATAATAAT 2755  
 QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
 Db 2756 GAATATTTCAATAGTATTAAATGATTTTGGAGCTTACAAAACAAAAAATGCT 2815  
 QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
 Db 2816 TTAGTGATACATCAGGATATATATCGAGAGTGAGGAGATGAGGAGATGTTCAACTTAAT 2875  
 QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
 Db 2876 ACCGATATACAAATGACCTTAAATTAAGTAGTTCAGAGAT-----AAAATTATA 2926  
 QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
 Db 2927 GTAAATTTAAATAATATATTTTATATAGCGCTATTTATGAGAGACTCTAGTGTAGTTT 2966



```

WKLEYISSNGPREFSVAENKVLAYNDGFIYVLSSSNSNIWNPIKTAINSYIICTL
SIVNVDYATWIDNNNNITDQPIINLPNPNFNNNSQIILKLEKL"
complement(3030..3905)
/gene="hai"
CDS
complement(3030..3905)
/gene="hai"
/gene="hai"
/codon_start=1
/transl_table=11
/product="HAI"
/protein_id="BA075082.1"
/db_xref="GI:4579736"
/transl_table="MCGNNSQTNANDLRNNEVFFISPSNNINKVLDKISQSEVKLNW
KLSGANKOBLIYDITNKQAVKIKVMDNTISLTNAPLSSVSQVCKITDNDNYWLLQ
NYISRVYIRVYNPNLQYINDIDTLMVSTOTSSNQPFKFSNKVICTDNLNRRNKLO
TOLNSDFLSKNSOILVILWOFDSROKWLIEYNETKSAYLLKQENRYLTION
SNVYEQYQSDLSLIQVWNTNYLDNASKYILYNLQDTRVLDVINSQIANGTHIVD
SIFGNTNQWMIINLI"
4153..7743
/gene="ntnh"
CDS
4153..7743
/gene="ntnh"
/codon_start=1
/transl_table=11
/product="NTNH"
/protein_id="BA075083.1"
/db_xref="GI:4579737"
/transl_table="MDINDOLINISVDMKNVIVVRKNTFFKAFKVPNIWVAPE
RYGEPDIAEYKLDGIVDSNFDQSDRENFLOAIILILKRINNTISGKOLLSLI
STAPFYGVGGYSSPNFTFGKPKSKNLSLVTSTIPFGGVRETNVIESON
NKNEYASNIVFGPSNIVENNVIYKKDAENGXGTMAEIVFPQLIYKYNKFIDP
AMETKLIISFLIGIKSDNLVYFRLTELDNKPQSQNLIDILISLGGVDEFI
NTPYFTNSIFPNSIKMFEKIKYIKYTEIEGNAIGNDIKLRKQKFOINQDILWN
NLNFCQSPNIIIDRFSSNAKHFYKQYITDNTDYNINGFVNGQINIKLPSLNK
TNIISKEKVVNLVNNENISLMSKNIYDGLGTEDFSTYKIPYNEEYEVFNDSO
NPLNNISIEBDSVPIPIIDIPYKONSNDLVFTQITSMTEVTHTALSINYLQAI
TNNENFTLSDESKVSSDKSVLYSFLDNLMSYLETINKDGPIDTKKYVILKEVF
KNYSFINDLQEDISGNEVWLPWPKALNLSYTSNVEYQDSGALSLSKDNL
REPNEIDDSLSGLKLNKLFYISKNIVYFKIYFSLDQWMTLSIYQYFE
LICKAKOSILQAQSLVQIVONKPTLDSKASIPPTKLIRETEKTFIDLSNESQIS
MNVNDFLNKASICVFVEDIYKPIESYMEKIYINIKRIFQICRTNIDNESKILI
NSYTKIDFEDFLIDIKNFNSQVQVMKELSPYQLLFASKSPNISIEDSGK
NLTIQYTESIELVGVNGESLSYKSPNETIKFSKFTNGLNNTFCWLSFTGKND
DKTLGNKNQKNWETYPEDNGLVPEIIDSNGQESVYLSNINDNYIISIVDRL
KQDILLINDKNQVANSIDOLISYTNIIISLVKNNSIVVELSVLNDNPITSEVIR
NYFSKLSNITRSKSLSEYKNYQLYNVEPETSLYEVDNNSKLSKNTDGINI
SSVFKLINDKESQYVQKKDECIIICVDGTEKYLQISPNRNRIQVLSKDNKAKTIV
NTDLFRPDCITFSYNDKIFSLSURDGDYNNWICNDNNKVPKGAHLWILES"
7758..11584
/gene="ntx"
CDS
7758..11584
/gene="ntx"
/codon_start=1
/transl_table=11
/product="NTX"
/protein_id="BA075084.1"
/db_xref="GI:4579738"
/transl_table="MTWPVKDFNYSDPVNDNDILYIRPONKLIITPPVKAFTINRDI
VIPRESDDNPSISKPRPTSKVQSYDPSVLTDEQKDTFLGKILKLRKINREDI
GKILNIVLGVSPMGSSPTDEDTHTTINIAVEKPEKNSKWNITLITPSVILRG
PLPNILDTASHTLQGGQSPSEFGFTSILKVAPEFLITLSDVTSNQSASVLAESI
FCMDPVALMHELTLHQYLGINIPSKRIPQVSEGFQSDGPNVQEEELYTFGGL
DVEIIPQERSQREKALGHYKDIARELNINKITPSSWISNDIKYKIFSEKYNFDK
DNTGNFVNDIKNSLYSDLTVMSEVYVSYQYNKRNTHYFSRHYLPVFAVILDDNI
YTIKRGNLNKNKINSQNTERNPALQKLSSEVVDLFTKVCURLTKNSRDSSTC
IKVIRNPLPVVADKDSISQIFENKIITDEFYQNTYSDKLSDELSDGQVFNPEIV
DPLIPVNMPELNPGEIYVYDITKYVDYLSYVLESQKLSNVMENITLTSVEE
ALGYSNKIIFPLSLAEKYNKGVQAGLFLANWEEVDEFTINMKKDTLKDSDSVI
IPYIPALNINSALRGNQAFATAGAVFLFEGFEFTIPALGVYDTSYSQADAKTIDL
IKTENCLEQKRWKQSDYQMWNSLRSITTOFNINTQWSDLSYQADAKTIDL
EYKYSQSDKENIKSOVENIKNSLDVKISEAMNNINKFIRECSVTYLFKNLPKVIDE
LNKEDLPRTKELILDSHNIILVGEVDRLKAKEVNSFENPMFNFYSFNLSLKDVI
INVENSTINDSKILSLOKNALVDTSYNAEVRVGDGNOLNTIYNDFKLSSGDKI
IVNLNNTLISAIYENSVSFWIKISKDLINSHNEVITIINSIQSGKLCIRNGNIE

```

```

BASE COUNT 4511 a 1227 c 1612 g 4234 t
ORIGIN
Alignment Scores:
Pred. No.: 2.68e-49 Length: 11584
Score: 867.50 Matches: 195
Percent Similarity: 61.64% Conservative: 83
Best Local Similarity: 43.24% Mismatches: 146
Query Match: 36.28% Indels: 27
DB: 7 Gaps: 11
US-09-910-186A-10 (1-450) x AB012112 (1-11584)
QY 2 ThrillePropheAsnillePheSerTyrThrAsnAsnSerLeuLeuLysAspIlelleAsn 21
|||||
Db 10272 ACAATGCGCTTTTAAATATTTTTCATATACATAAATCTTTTATAAAGATATAATTAAT 10331
QY 22 GluTyrPheAsnAsnilleAsnAspSerLysLysLeuSerLeuGlnAsnArgLysAsnThr 41
|||||
Db 10332 CAATATTTTCAATAGTATTAATGATTCATAAATTTTGAGCTTACAAAACAAAAAATGCT 10391
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
|||||
Db 10392 TTAGTGGATACATCAGGATATAATGCAGAGTGGAGGTAGAGATAATGTTCACTTAAT 10451
QY 62 ProillePhePropheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValille 81
|||||
Db 10452 ACGATATATACAAATGACTTTAAATTAAGTAGTTACAGAGAT-----AAAAATATA 10502
QY 82 ValThrGlnAsnGluAsnilleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
|||||
Db 10503 GTAAATTTAAATAATAATATTTTATATAGCGCTATTTATAGAACTCTAGTGTAGTTT 10562
QY 102 TrpilleArgilleAsnLysTrpValSerAsnLeuProgly---TyrThrilleAspSer 120
|||||
Db 10563 TGGATTAAGATATCTAAAGATTTAACTAAATTTCTCATATGAATATAAATAAATACAGT 10622
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleSerAsnPheLeuValPheThrLeu 140
|||||
Db 10623 ATAGAACAAATTTCTGGTGGAAATTTATGATTAGGAATGCCAATATAGATGATTTTA 10682
QY 141 LysGlnAsnGluAspSerSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
|||||
Db 10683 CAAGATGTTAATAGAAAGTATAAAGTTTAAATTTTGTATTATAGTGAATCATTAAGTCAT 10742
QY 161 ProglyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMetGlyAsnMet 179
|||||
Db 10743 ACAGATATACAAATAAATAGTTTGTGTTACTATACTATAATAATAATAGGGGTATATG 10802
QY 180 LysilleTyrilleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199
|||||
Db 10803 AAACCTTTATATAAATGGAGAATTAAGCAGAGCTCAAAAATTTGAAGATTTAGATGAGTT 10862
QY 200 AsnPheSerLysThrillePheGluIleAsnLysIleProAspThrGlyLeuIleThr 219
|||||
Db 10863 AAGTTAGATAAACCACATAGTATTGGGATAGATAGATAATAGAT----- 10907
QY 220 SerAspSerAsnilleAsnMetTrpIleArgAspPheThrillePheAlaLysGluLeu 239
|||||
Db 10908 -----GAGAATCAGATCTTTGGATTAGACATTTAATATTTTCTAAGAATA 10958
QY 240 AspGlyLysAspIleAsnilleLeuPheAsnSerLeuGlnThrThrAsnValLysAsp 259
|||||
Db 10959 AGTAATAGAGATATTAATTTGTTATATAGGACAAAATATTAAAGAAATGTTATAAAGAT 11018
QY 260 TyrTTPGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279
|||||

```



[illegible]

Search completed: November 7, 2002, 18:38:07  
Job time : 2772 secs

Copyright from 1 09 10 to 86 and 05-05-2004

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 16:30:36 ; Search time 253 seconds  
(without alignments)  
4005.531 Million cell updates/sec

Title: US-09-910-186a-10

Perfect score: 2391

Sequence: 1 MTIFNFISVTNNSLLKDI.....NYASLLESTHMGFVPSV 450

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

MODEL=frame\_p2n.model -DEV=xlh  
-c/cn2\_1/USPro\_spool/US09910186/runat\_04112002\_111616\_20826/app\_query.fasta\_1.647  
-N=N\_Geneseq\_101002 -OFT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-COOEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -NCPEN=2000000000  
USER=US09910186 -CGEN.1.1.125 -runat\_04112002\_111616\_20826 -ICPU=3  
-NO\_XLIPY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMESOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002.\*

1: /SID82/cgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID82/cgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID82/cgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID82/cgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID82/cgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID82/cgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID82/cgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID82/cgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID82/cgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID82/cgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID82/cgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID82/cgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID82/cgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID82/cgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID82/cgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID82/cgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID82/cgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID82/cgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID82/cgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID82/cgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID82/cgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID82/cgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID82/cgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID82/cgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2391	100.0	1371	21	AA287214
2	2391	100.0	1371	22	AAA54486
3	2386	99.8	3950	20	AA25521
4	2319	97.0	1502	19	AAV30588
5	1589	66.5	1208	21	AA298630
6	1589	66.5	1440	21	AA298631
7	868.5	36.3	1374	22	AAA54487
8	803.5	33.6	1469	19	AAV30591
9	583.5	24.7	1400	21	AA287215
10	589.5	24.7	1400	22	AAA54488
11	586.5	24.5	2532	21	AA287218
12	586.5	24.5	3891	17	AA292944
13	586.5	24.5	4835	21	AA264582
14	583	24.4	4017	22	AA298632
15	581	24.3	3876	19	AAV30579
16	578.5	24.2	1472	19	AAV30585
17	576	24.1	1463	19	AAV30584
18	575.5	24.1	1347	21	AA287217
19	575.5	24.1	1368	22	AA254491
20	575.5	24.1	1535	19	AAV30586
21	565	23.6	1526	19	AAV30580
22	558.5	23.4	1402	17	AA292946
23	558.5	23.4	1402	19	AAV30572
24	558.5	23.4	1546	19	AAV30575
25	557.5	23.3	1317	21	AA287220
26	557.5	23.3	1330	17	AA292945
27	557.5	23.3	1330	19	AAV30571
28	557.5	23.3	1332	22	AAA54482
29	557.5	23.3	1338	21	AA287216
30	557.5	23.3	1351	19	AAV30576
31	556.5	23.3	1323	22	AAA54483
32	556.5	23.3	1326	22	AAA54484
33	553.5	23.0	1278	22	AA254489
34	550	23.0	1547	19	AAV30581
35	546	22.8	1293	18	AA148100
36	546	22.8	1313	18	AA148101
37	546	22.8	1314	22	AAA54499
38	546	22.8	1317	22	AAA54490
39	542	22.7	1341	21	AA287213
40	542	22.7	1341	22	AAA54485
41	539.5	22.6	3509	19	AAV26289
42	531	22.2	1317	21	AA287216
43	511.5	21.4	1460	19	AAV30593
44	492	20.6	1347	22	AAA54589
45	471	19.7	1332	22	AAA54588

ALIGNMENTS

RESULT 1  
AA287214  
ID AA287214 standard; DNA; 1371 BP.  
XX  
XX  
XX  
XX  
DT 08-MAY-2000 (first entry)  
XX  
DE DNA encoding synthetic BoNT serotype C (BoNTC) Hc fragment.  
XX  
KW Botulinum neurotoxin; heavy chain; BoNT; serotype C;  
C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
KW VEE; botulinism; vaccine; diagnosis; drug screening; ds.  
XX  
XX Clostridium botulinum.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers

FT CDS 10..1362  
 FT /\*tag= a  
 FT /product= "Synthetic botulinum neurotoxin serotype C  
 FT (BoNTC) heavy chain C-terminal fragment (Hc)"  
 PN WO200002524-A2.  
 XX 20-JAN-2000.  
 XX 09-JUL-1999; 99WO-US15570.  
 XX 10-JUL-1998; 98US-0092416.  
 XX 12-MAY-1999; 99US-0133870.  
 XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
 XX

Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh WT, Smith L;

WPI; 2000-160827/14.

P-PSDB; RAY77136.

Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum toxin serotypes A-G, is used for inducing an immune response against botulinum.

Disclosure; Page 41-42; 54pp; English.

The invention relates to novel vaccines that induce a protective immune response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant DNA construct comprising a vector, and at least one nucleic acid fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-G. In preferred embodiments of the invention, the vector is a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this vector results in the production of large amounts of a protein encoded by a sequence cloned into the replicon. The constructs are used to produce vaccines against botulism. The proteins can also be used as diagnostic tools for the diagnosis of botulism. The transformed host cells can be used to analyse the effectiveness of drugs and agents which inhibit toxin effects. The vaccine currently used against botulism is dangerous and expensive to produce, and contains formalin, which is very painful for the recipient. Also, the vaccine is incomplete, in that only 5 of the 7 serotypes are represented in the formulation. The novel vaccine of overcomes these problems, as it is easily purified, and available in large quantities. It is also expressed in the lymph nodes for a better immune response. Sequences AAZ87212-287217 represent synthetic DNA sequences encoding BoNT Hc fragments used in the present invention. These were optimised for codon usage for expression in yeast.

Sequence 1371 BP; 397 A; 416 C; 265 G; 293 T; 0 other;

#### Alignment Scores:

Seq. No.:	Length:	1371
Score:	2391.00	450
Percent Similarity:	100.00%	0
Best Local Similarity:	100.00%	0
Query Match:	100.00%	0
DB:	21	0

US-09-910-186a-10 (1-450) x AAZ87214 (1-1371)

QY 1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIle 20  
 DB 10 ATGACCATTCCATTCATCTCTCTACACCAACAACCTCCCTGTGTGAAGGACATCATC 69  
 QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40  
 DB 70 AACGAGTACTTCAACAACATCAAGACTTCCAGACTCTGCTCCCTGCGAGACCTAGAAC 129  
 QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60  
 DB 130 ACCTTGGTCGACACCTCCGGTGTACAAACCGGAGGTCTCCGAGGAGGGGTGACGTCACGCTG 189

QY 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80  
 DB 190 AACCAATCTCCCATTCGACTTCAAGCTGGGTCTCCGGTGAGGACAGAGTAAAGTTC 249  
 QY 81 IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100  
 DB 250 ATCGTCAACCCAGAACGAGAACATCGTCAACACTCCATGTACGAGTCTTCTCCATCTCC 309  
 QY 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120  
 DB 310 TTCGTGATCAGAATCAACAAGTGGTCTCAACTGCCAGGTTACACCATCATCGACTCC 369  
 QY 121 ValLysAsnAsnSerGlyTyrPheIleGlyIleSerAsnPheLeuValPheThrLeu 140  
 DB 370 GTCAGAACAACTCCGGTGTGTCATCGGTATCATCTCCAACCTCTCTGCTCTTCCACCTG 429  
 QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAla 160  
 DB 430 AAGCAGAACGAGGACTCCGAGCAGTCCATCACTTCTCTACGACATCTCCAACACGCT 489  
 QY 161 ProGlyTyrAsnLysTrpPheValThrValThrAsnAsnMetMetGlyAsnMetLys 180  
 DB 490 CCGTGTACAAACAGTGGTCTTCCGTACCGTCAACCAACATGATGGTAAACATGAAG 549  
 QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200  
 DB 550 ATCTACATCAACGGTAAGCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAAC 609  
 QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220  
 DB 610 TTCCTCCAGACCATCCTTCGAGATCAACAAGATCCAGACACCGGTCTGTATCACTCC 669  
 QY 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240  
 DB 670 GACTCCGACAAACATCAACATGTGATCGGTGACTTCTATCATCTTCGCCAAGGAGTTGGAC 729  
 QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlyTyrThrAsnValValLysAspTyr 260  
 DB 730 GGTAAAGACATCAACATCCCTGTTCACCTCTTCGAGTACACCAACCTCGTCAAGGACTAC 789  
 QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGlyTyrMetValAsnIleAspTyrLeuAsn 280  
 DB 790 TGGGTTAACGACCTCGAGATACACAGAGGAGTACTATAGTGTCAACATCGACTACTTTGAAC 849  
 QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsp 300  
 DB 850 AGATACATGTACGCCAATCTCGACAGATCGTCTTCAACACACGAGCTAACACACACGAC 909  
 QY 301 PheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320  
 DB 910 TTCAACGAGGGTTACAAGATCATCATCAAGCGTATCAGAGGTAACACCAACGACACACAGA 969  
 QY 321 ValArgGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340  
 DB 970 GTCAGAGGTGTGACATCCCTGACTTTCGACATGACTATCAACAACAGGCTTACACCTTG 1029  
 QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
 DB 1030 TTCATGAAGAACGAGACCATGTACGCCGCAACACCACTCCACCGGAGGACATCTACGCCATC 1089  
 QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380  
 DB 1090 GGTCTGGCTGAGCAGACCAAGGACATCAACGACAAACATCATCTTCCAGATCCAGCCAATG 1149  
 QY 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400  
 DB 1150 AACACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACCTCAACGGTGAGAACATC 1209  
 QY 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHis 420  
 DB 1210 TCCGGTATCTTCCATCGGTACCTACAGATTCGGTCTGGGTGGTGGTGGTACGACAC 1269  
 QY 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440

Db 1270 AACTACTTGGTCCCAACGTCAAGCAGGTAACTACGCTCTTGTGAGTCCACTTCC 1329  
 |||

QY 441 ThrHisTrpGlyPheValProValSerGlu 450  
 |||

Db 1330 ACCCATGGGGATTGCTCCAGTCCCGAG 1359  
 |||

RESULT 2

AAA54486

ID AAA54486 standard; DNA; 1371 BP.

XX AC

XX AC

11-APR-2001 (first entry)

DE Botulinum toxin heavy chain C-terminal coding sequence (serotype C).

Botulinum toxin; neurotoxin; heavy chain; recombinant expression;  
 recombinant vector; antigen; immune response; vaccine; bacterium;  
 infection; ds.

PS Synthetic.

PS Clostridium botulinum.

Key Location/Qualifiers

CDS 10..1362

/\*tag= a

/product= H\_C peptide fragment

WC2000067700-A2.

16-NOV-2000.

12-MAY-2000; 2000WO-US12890.

12-MAY-1999; 99US-0133865.

12-MAY-1999; 99US-0133865.

12-MAY-1999; 99US-0133867.

12-MAY-1999; 99US-0133868.

12-MAY-1999; 99US-0133869.

12-MAY-1999; 99US-0133873.

29-JUL-1999; 99US-0146192.

(USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

WPI; 2001-016048/02.

P-PSDB; AAB04092.

New nucleic acids encoding the carboxy- or amino-terminal portions of  
 the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 vaccine against botulism

Claim 2; Fig 5a; 73pp; English.

Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
 chain and then posttranslationally nicked, forming a dichain  
 consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 remain linked by a disulfide bond. Nucleic acids encoding the  
 carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 expression vectors and expressed in transformed cells to produce  
 peptide antigens useful for eliciting an immune response to give  
 protective immunity against botulinum neurotoxin, which causes  
 botulism. The nucleic acids are expressible in a recombinant  
 organism such as *Escherichia coli* or *Pichia pastoris*. The use  
 of recombinant nucleic acids are advantageous since it eliminates  
 the need to culture large quantities of hazardous toxin-producing  
 bacterium. Production yield from the genetically engineered product  
 is also high and cost of production is lower. The nucleic acids can  
 be derived from clostridium botulinum serotypes A-G.

XX

Alignment Scores:

Score: 1.42e-205 Length: 1371  
 Pred. No.: 2391.00 Matches: 450  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0

US-09-910-186A-10 (1-450) x AAA54486 (1-1371)

QY 1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleile 20  
 |||  
 Db 10 ARGACCATCCCATTCACACATCTTCTCCACCAACACTCCCTGTTGAAGGACATATC 69  
 |||  
 QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40  
 |||  
 Db 70 AACGAGTACTTCAACACATCAACAGACTCCAAAGATCTGTCCTCCAGAACCGTAAGAAC 129  
 |||  
 QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60  
 |||  
 Db 130 ACCTTGGTCGACACCTCCGGTTACAACGCCGAGGTCTCCGAGGAGGTGACGTCACGCTG 189  
 |||  
 QY 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80  
 |||  
 Db 190 AACCCAATCTTCCCATTCGACTTCAAGCTGGTCTCCCGGTGAGGACAGAGTAAGGTC 249  
 |||  
 QY 81 IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100  
 |||  
 Db 250 ATCGTCACCCAGACGAGACATCGTCTACCACTCCATGACGAGTCCCTTCCATCTCC 309  
 |||  
 QY 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120  
 |||  
 Db 310 TTCTGATCAGAATCAACAAGTGGTCTCCAACTTGCCAGGTTACACCATCATCGACTCC 369  
 |||  
 QY 121 ValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140  
 |||  
 Db 370 GTCAAGAACAACCTCCGGTTGGTCCATCGGTATCATCTCCAACCTTCTGGTCTTACCCCTG 429  
 |||  
 QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160  
 |||  
 Db 430 AAGCAGAAGGAGGACTCCGACGACTCCATCAACTTCTCTACGACATCTCCACAACGCT 489  
 |||  
 QY 161 ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180  
 |||  
 Db 490 CCTGGTTACAACAAGTGGTCTTCTGTCACCGTCCACCAACATGATGGTAAACATGAAG 549  
 |||  
 QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200  
 |||  
 Db 550 ATCTACATCAACGGTAACTGATCGACACCATCAAGTCAAGGAGTTGACGGGTATCAAC 609  
 |||  
 QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220  
 |||  
 Db 610 TTCTCCAAGACCATCACCTTCGAGATCAACAAGATCCACGACACCGGTCTGATCACTCC 669  
 |||  
 QY 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240  
 |||  
 Db 670 GACTCCGACACATCAACATGTGGATCGGTACACTTCTACATCTTCGCCCAAGGAGTTGGAC 729  
 |||  
 QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyr 260  
 |||  
 Db 730 GGTAAGGACATCAACATCTCTTCAACTCTCTTGCAGTACACCAACGTCGTCAGGACTAC 789  
 |||  
 QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTrpMetValAsnIleAspTyrLeuAsn 280  
 |||  
 Db 790 TGGGTAAAGACCTGAGATACAAAGGAGTACTACATGGTCAACATCGCTACTTGAAC 849  
 |||  
 QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsp 300  
 |||  
 Db 850 AGATACATGTAGCCCAACTCCACAGACAGATCTCTTCAACACCAACGACGTAACACACGAC 909  
 |||  
 QY 301 PheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320  
 |||



|||||  
Dd 910 TTCACAGGGTTACAGATCATCATCAACCGTATCAGGATACACCAACACACACCA 969  
Qy 321 ValArgGlyAspIleLeuTyrPheAspMetThrIleAsnLysAlaTyrAsnLeu 340  
Dd 970 CTCAGAGTGGTACATCCCTGACTTCGACATGACATCAACAACAGCCCTACAACCTG 1029  
Qy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
Dd 1030 TTCATGAGACGAGACCATGTACGCGCAACCACTCCACGAGGACATCACCACATC 1089  
Qy 351 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIlePheGlnIleGlnPromet 380  
Dd 1090 GGTCTGGTGGACGAGACCAAGGACATCAAGCAACATCATCTCCAGATCCAGCAATG 1149  
Qy 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400  
Dd 1150 AACACACTTACTACTAGCTTCCAGATCTTCAAGTCCACTTCAACGGTGAGAACATC 1209  
Qy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyAspTrpTyrArgHis 420  
Dd 1210 TCCGGTATCTGTTCCATCGGTACTACAGATCCGTCGTGGTGACTGCTACAGACAC 1269  
Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440  
Dd 1270 AACACTTGGTTCCAACTGTCAAGCAGGTAACACTACGCTCTCTGCTGGAGTCCACTTCC 1329  
Qy 441 ThrHisTrpGlyPheValProValSerGlu 450  
Dd 1330 ACCCACTGGGATTCGTCGCCAGTCTCCGAG 1359

RESULT 3  
25521  
AAAX25521 standard; DNA: 3950 BP.  
AAAX25521;  
02-AUG-1999 (first entry)  
DNA coding for modified botulinum toxin rBont/C.

Botulinum toxin; botulism; rBont/C; vaccine; drug delivery;  
mutant; ss.

Clostridium botulinum.  
Synthetic.

Key Location/Qualifiers  
CDS 32..3907  
/\*tag= a

WO9920306-AL.

29-APR-1999.

16-OCT-1998; 98WO-US21897.

20-OCT-1997; 97US-0954302.

(UJJE-) UNIV JEFFERSON THOMAS.

Kiyatkin N, Maksymowych A, Simpson L;

WPI; 1999-302646/25.

P-PSDs; AAY05814.

Modified toxin useful for systemic delivery of oral vaccines and

therapeutic agents

Example 1; Page 26-31; 37pp; English.

This DNA sequence codes for a modified serotype C botulinum toxin,

termed rBont/C (see AAY05814), in which amino acids His-229, Glu-230

CC and His-233 of the native sequence are substituted by Gly, Thr and  
CC Asn, respectively, i.e. the zinc binding motif (see AAY05817) of  
CC the light chain holotoxin is modified, resulting in loss of  
CC endoprotease activity. DNA coding for the modified botulinum toxin  
CC was assembled from 3 separate toxin fragments using PCR and  
CC site-directed mutagenesis. The modified recombinant botulinum  
CC toxin maintains its ability to translocate from the gut into the  
CC general circulation but is non-toxic. It can be used as an oral  
CC vaccine for antigenic peptides including botulinum toxin (i.e. an  
CC oral vaccine for botulism) or for the oral delivery of other  
CC therapeutic agents to the general circulation.

XX  
SQ Sequence 3950 BP; 1589 A; 410 C; 616 G; 1335 T; 0 other;

Alignment Scores:

Prod. No.: 1.43e-204 Length: 3950  
Score: 2386.00 Matches: 449  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.79% Indels: 0  
DB: 20 Gaps: 0

US-09-910-186A-10 (1-450) x AAX25521 (1-3950)

Qy 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeu 21  
Dd 2558 ACAATACCCCTTAAATATTTTTCATATACTAAATAATCTTTTAAAGATATAATTAAT 2617  
Qy 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
Dd 2618 GAATATTTCATTAATATTAATGATTCAAAATTTTCAGCCTACAAAACAGAAAATACT 2677  
Qy 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
Dd 2678 TTACTGGATACATCAGGATATAATGCAAGAGTGAAGTGAAGAGGCGATGTTTCAGCTTAAT 2737  
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
Dd 2738 CCAATATTTCCATTTGNCITTAATTAATAGGTAGTTTCAGGGAGGATAGAGTAAAGTTATA 2797  
Qy 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
Dd 2798 GTAACCCAGAAATGAAATATTTGTATATAATTCATGTATGATAAGTTTATAGCATTAGTTT 2857  
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
Dd 2858 TGGATTAGAATAAATAATGGTAAGTAATTTACCTCGATATACATAATTCATATGATAGTGT 2917  
Qy 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
Dd 2918 AAAAATAACTCAGGTGAGTATAGGTATATTAGTAATTTTTCATATTACTTTAAAA 2977  
Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161  
Dd 2978 CAAATGAAGATAGTGAACAAAGTAAATTTTAGTTATGATATATCAAAATATGCTCTCT 3037  
Qy 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
Dd 3038 GGATACATAAATGGTTTTTTTGTAACTGTTACTAACAAATATGATGGAAATATGAAGATT 3097  
Qy 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
Dd 3098 TATATAATGGAAATTAATAGATACTATAAAAAGTTAAAGAACTAACTGGAATTAATTTT 3157  
Qy 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221  
Dd 3158 AGCAAACTATACATTTGAATAAATAAATAATTCAGATACCGGTTTGATTACTTCCAGAT 3217  
Qy 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241  
Dd 3218 TCTGATACATCAATATGCGATAAGAGATTTTATATATTTTCTGCTAAAGAAATAGATGCT 3277  
Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261

Db 3278 AAAGATATTAAATATATTATTAATAGCTTGAATATATACTAATATGTTGAAAAAGATTATTCG 3337  
 QY 262 GIVASnAspLeuArgTyrAsnLysGluTyrMetValAsnIleAspTyrLeuAsnArg 281  
 Db 3338 GGAATGATTTAAGATATAAATAAGATATTATATGGTTAATATAGATTATTTAAATAGA 3397  
 QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
 Db 3398 TATATGATGCGAACTCAGCAAAATTCCTTTTAAATACACGTAGAAATATAATAGACTTC 3457  
 QY 302 AsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
 Db 3458 AATGAGGAGTATAAATATATAAATAAGATCAGAGAAATACAATATGATAGTAGTA 3517  
 QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
 Db 3518 CGAGGAGGAGATATTATTTATGATGACATTAATAACAAGCATATAATTTGTTT 3577  
 QY 342 MetLysAsnGluThrMetTyrAlaAsnAsnHisSerThrGluAspIleTyrAlaIleGly 361  
 Db 3578 ATGAAGAATGAACATATGATGACAGATATCATAGTACTGAAGATATATATGCTATAGGT 3637  
 QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleLeuPheGlnIleGlnProMetAsn 381  
 Db 3638 TTAAGAGAACAAACAAGGATATAAATGATATATATTTTCAATACACCAATGAAT 3697  
 QY 382 AsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
 Db 3698 AATACTATTATATAGCATCTCAATATATTAATCAAAATTTTAAATGGAGAAATATTTCT 3757  
 QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsn 421  
 Db 3758 GGAATATGTTCAATAGGTACTATGCTTTAGACTGGAGGTGATTTGATAGACACAAAT 3817  
 QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441  
 Db 3818 TATTTGGTCCCTACTGTGAAGCAAGGAAATATGCTTCATTATTAGAATCAACATCAACT 3877  
 QY 442 HisTrpGlyPheValProValSerGlu 450  
 Db 3878 CATGGGGTTTGTACCTGTAGTGAA 3904  
 RESULT 4  
 AAV30588  
 AAV30588 standard; DNA; 1502 BP.  
 AAV30588;  
 07-DEC-1998 (first entry)  
 Clostridium botulinum type C1 toxin C fragment gene in pETHisb.  
 Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;  
 botulism; Botc; ds.  
 Clostridium botulinum serotype C1 Stockholm strain.  
 Synthetic.  
 Key Location/Qualifiers  
 CDS 108..1463  
 /\*tag= a  
 WO9808540-A1.  
 05-MAR-1998.  
 28-AUG-1997; 97WO-US15394.  
 28-AUG-1996; 96US-0704159.  
 (OPHI-) OPHIDIAN PHARM INC.

PI Thalley BS, Williams JA;  
 XX WPI: 1998-230234/20.  
 DR P-PSDB; AAW68397.  
 XX Host cell containing recombinant expression vector encoding  
 PT Clostridium botulinum type B or E toxin - useful to treat humans  
 PT and other animals at risk of intoxication with clostridial toxin  
 XX Example 46; Page 339-341; 428pp; English.  
 XX This is the DNA sequence of the Clostridium botulinum serotype C1  
 CC (Stockholm strain) neurotoxin fragment C gene contained in plasmid  
 CC pETHisb. The encoded BotC fragment C polypeptide (see AAW68397) has a  
 CC His-tagged N-terminal extension. The vector was used to express  
 CC native (i.e. non-fusion) soluble C fragment in Escherichia coli host  
 CC cells. The invention relates to recombinant proteins derived from  
 CC C. botulinum toxins, especially type B and type E toxins. Methods  
 CC are provided which allow for the isolation of soluble recombinant  
 CC proteins free of significant endotoxin contamination. Preferred  
 CC hosts for production of recombinant proteins are E. coli, insect  
 CC cells and yeast cells. The recombinant toxins are used as  
 CC immunogens for the production of vaccines and antitoxins that are  
 CC useful in the treatment of humans and animals at risk of  
 CC intoxication with clostridial toxin.  
 XX Sequence 1502 BP; 600 A; 155 C; 249 G; 498 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 4.67e-199 Length: 1502  
 Score: 2319.00 Matches: 439  
 Percent Similarity: 98.66% Conservative: 3  
 Best Local Similarity: 97.99% Mismatches: 5  
 Query Match: 96.99% Indels: 1  
 DB: 19 Gaps: 0  
 US-09-910-186A-10 (1-450) x AAV30588 (1-1502)  
 QY 4 PropHeAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleLeuAsnGluTyr 23  
 Db 152 CCATATCGAAGTCTCATATGCTAGCTAGCTTATTAAAGATATATTAATGAATA 211  
 QY 23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43  
 Db 212 TTTCAATATATATATGATTCAGGCTTCAAAATTTTGGAGCTCAAAACAGAAAAATCTTAGT 271  
 QY 43 lAspThrSerGlyTyrAsnAlaGluValSerGluGlyAspValGlnLeuAsnProIle 63  
 Db 272 GGATACATCAGATATATATGAGAGTGAGTGAGAGAGGCGATGTTGAGCTTAATCCAAT 331  
 QY 63 ePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleValTh 83  
 Db 332 ATTTCATTTCACCTTAAATTAGGTAGTTTCAGGGGAGGATAGAGTAAAGTTATAGTAAC 391  
 QY 83 rGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrpIle 103  
 Db 392 CCAGAATGAATATATGATATATTTCTATGATCAAAAGTTTACGATTTAGTTTGGAT 451  
 QY 103 eArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerValLysAs 123  
 Db 452 TAGAATAAATAAATGGTAAAGTAAATTTACCTGGATATATACTATAATGATAGTAAAAA 511  
 QY 123 nAsnSerGlyTrpSerIleGlyIleIleSerAsnPheIleValPheThrLeuLysGlnAs 143  
 Db 512 TAACTCAGTTGGAGTAGGTATATATTAGTAATTTTATTAGTATTTTAAACAAAAA 571  
 QY 143 nGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTy 163  
 Db 572 TCAAGATAGTGAACAAAGTAAATTTTAGTTATGATATATCAAAATATGCTCTCGATA 631  
 QY 163 rAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIle 183  
 Db 632 CAATAAATGGTTTTTTTGTAACTGTACTTAACAATATGATGGGAAATATGAAGATTATAT 691

QY 183 eAsnGlyLysLeuLeuAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLys 203  
 DB 692 AAATGGAAATTAATAGACTATATAAAGTAAAGAACTAACTGGAATTAATTTAGCAA 751  
 QY 203 sThrIleThrPheGluLeuLeuLysIleProAspThrGlyLeuIleThrSerAspSerAs 223  
 DB 752 AACTATAACATTTGAATATAAATAATCCAGATACCGTTTGATTCAGATTCTGA 811  
 QY 223 pAsnIleAsnMetTrpIleArgAspPheTrpIlePheAlaLysGluLeuAspGlyLysAs 243  
 DB 812 TAACATCAATATGCGATAACAGATTTTATATATTTGCTAAAGAAATTAGATGTAAGA 871  
 QY 243 pIleAsnIleLeuPheAsnSerLeuGlnThrThrAsnValValLysAspTrpIleGlyAs 263  
 DB 872 TATTAATATATTTTAACTGTCGCAATATGTTTAAAGATGTTTGGGAA 931  
 QY 263 uAspLeuArgTrpAsnLysGluTrpTrpMetValAsnIleAspTrpLeuAsnArgTrpMe 283  
 DB 932 TGATTTAAGATATAATAAAGAAATATATGTTTAAATAGATTTTAAATAGATATAT 991  
 QY 283 tTyralAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnG 303  
 DB 992 GATCGCAACTCAGACAAATGTTTAAATACACGTAGAAATTAATGACTTCAATGA 1051  
 QY 303 uGlyTrpIleLeuIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgG 323  
 DB 1052 AGCATATAAATATATAAAGAAATCAGAGAAATACAAATGATACAGTACAGAGG 1111  
 QY 323 yGlyAspIleLeuTrpPheAspMetThrIleAsnAsnLysAlaTrpAsnLeuPheMetLys 343  
 DB 1112 AGGAGATATTTTATATTTGATGACAAATTAATAAAGACATATAATTTGTTATGAA 1171  
 QY 343 sAsnGluThrMetTrpAlaAspAsnHisSerThrGluAspIleTrpAlaIleGlyLeuAr 363  
 DB 1172 GAATGAACATGATGATGAGATATATCATCTACGACAGATATATATGCTATAGTTTAA 1231  
 QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleClnProMetAsnAsnTh 383  
 DB 1232 AGAACAAACAAAGGATATAAATGATATATATATATATTTCAATACAAACCAATGAATAAC 1291  
 QY 383 rTyrrTyrrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyI 403  
 DB 1292 TTATATTATACCACTCCAAATATTAAATCAATTTTAAATGGAGAAATATTTCTGGAA 1351  
 QY 403 eCysSerIleGlyThrTrpArgPheArgLeuGlyGlyAspTrpTrpArgHisAsnTrpLe 423  
 DB 1352 ATGTTCAATAGTACTTATCGTTTACACTTGGAGTGATTTGGTATAGACACAATATTT 1411  
 QY 423 uValProThrValLysGlnGlyAsnTrpAlaSerLeuLeuGluSerThrSerThrHisTr 443  
 DB 1412 GTGGCTACTGTGAAGCAAGGAATTTGCTTCATTTAGATCAACATCAACTCATTTG 1471  
 QY 443 pGlyPheValProValSerGlu 450  
 DB 1472 GGGTTTGTACCTGTAAGTGAA 1493

## RESULT 5

AAZ98630

ID AAZ98630 standard; DNA; 1208 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

Copied from 10910186 on 05-09-2004

XX 03-FEB-2000.  
 XX 20-JUL-1999; 99WO-IB01301.  
 XX 22-JUL-1998; 982A-0006538.  
 XX (AGRI-) AGRIC RES COUNCIL.  
 XX De Bruyn EE, Botha AD;  
 XX WPI: 2000-205375/18.  
 XX P-PSDB; AAY78982.

Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin type D, useful in vaccines for protection against botulism, comprises at least one amino acid mutation not present in the wild type D neurotoxins

Claim 8; Page 54-57; 66pp; English.

This sequence represents a nucleotide sequence of a synthetic gene encoding a non-toxic immunogenic derivative of Clostridium botulinum type D toxin (BoNT). Botulinum neurotoxin causes botulism poisoning in cattle and sheep, and usually results in the death of the affected or poisoned animal. The non-toxic immunogenic fragments of the C. botulinum neurotoxin are useful in vaccines to protect animals (e.g. humans, cattle, sheep, pigs) against BoNT type D poisoning. The non-toxic fragments can be produced relatively simply and inexpensively (specifically by fermentation techniques). As the fragments are not toxic the risk to production staff is reduced.

Sequence 1208 BP; 416 A; 203 C; 216 G; 373 T; 0 other;

## Alignment Scores:

Pred. No.: 1,43e-133 Length: 1208  
 Score: 1589.00 Matches: 307  
 Percent Similarity: 84.16% Conservative: 33  
 Best Local Similarity: 75.99% Mismatches: 56  
 Query Match: 66.46% Indels: 8  
 Gaps: 4

US-09-910-186A-10 (1-450) x AAZ98630 (1-1208)

QY 50 AlaGluValSerGluGluGlyAspValClnLeuAsnProIlePheProPheAspPheLys 69  
 DB 4 GCAGAAGTTAGAGTTGAGGGAACGTCAGTGAATCCTATCTTCATTGACTTTAAG 63  
 QY 70 LeuGlySerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleVal 89  
 DB 64 TTGGTAGCTCTGCTGATGACCGTGGCAAAATATTGTAACCAAAACGAAATATTGT 123  
 QY 90 TyrAsnSerMetTrpGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpVal 109  
 DB 124 TATAATGCTATGATGATCTTCTCTATTAGCTTCTGGATCAGATTAATAAATGGGT 183  
 QY 110 SerAsnLeuProGlyTrpThrIleIleAspSerValLysAsnAsnSerGlyTrpSerIle 129  
 DB 184 TCATATTACAGGTTATACCATCATCGACAGCGTTAAATAATATTCCAGATGCTATT 243  
 QY 130 GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer 149  
 DB 244 GGCATAATTAGCAATTTTCTAGTTTTTACACTAAAGCAAAACGAAATACCGACGAG 303  
 QY 150 IleAsnPheSerTrpAspIleSerAsnAsnAlaProGlyTrpAsnLysTrpPhePheVal 169  
 DB 304 ATTAATTTTCTTATGACATTTTCAAGAACGCTGCTGTTTATAATAATGGTCTTTGTA 363  
 QY 170 ThrValThrAsnAsnMetMetGlyAsnMetLysIleTrpIleAsnGlyLysLeuIleAsp 189  
 DB 364 ACCATTACCAACACATGATGGGTAATATGATGATCTACATTACGGAACGTAATCGAC 423  
 QY 190 ThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle 209

```

Db 424 ACCATCAAGTGAAGAAATGACCGGTATCACTTCTCAAGACCATACGTTTCAGATG 483
Qy 210 AsnlylleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetIle 229
Db 484 AATAAAATCCGACACAGAGGATTAATACGCTCTGATCTGACACATCATATGTGAT 543
Qy 230 ArgaspPheTyrllePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn 249
Db 544 AGAGATTCTATATTTTCGCAAAAGATTAAGACAGACATTAATATCCCTTTCAT 603
Qy 250 SerLeuGlnTyThrAsnValVallyAspTyTrpGlyAsnAspLeuArgTyAsnLys 269
Db 604 TCATCTCAGTACACTAATGTTGTGAAGCAATATTTGGGTAATGATTAAGATGATAAG 563
Qy 270 GluTyTyMetValAsnIleAspTyLeuAsnArgTyMetTyAlaAsnSerArgGln 289
Db 664 GAGTACTACATGATTAATGTAATATCATGATGATGATGATGATGATGATGATGAT 723
Qy 290 IleValPheAsnThrArgAsnAsnAspPheAsnGluGlyTyTyLysIleIle 309
Db 724 ATCGTCTTCAACACTCGTAAATAAACAAGATTTCAACAGGGTACAAAATATCAT 783
Qy 310 LysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyAspIleLeuTyPhe 329
Db 784 AAAAGATCCGTGGAATACAAATGATAGAGTTCGCTGGTGAGAACGTCATATTT 843
Qy 330 AspMetThrIleAsnAsnLysAlaTyAsnLeuPheMetLysAsnGluThrMetTyAla 349
Db 844 AATAACAGTAGATAATAACAGTCTCTTA-----GCCATGTATAA 888
Qy 350 AspAsnHisSerThr---GluAspIleTyAlaIleGlyLeuArgGluGlnThrLysAsp 368
Db 889 CGGTCAAGAACTTAGGACAGACCTTTGTCATTTGGTGCATGATCAACCGATGGAC 948
Qy 369 ----IleAsnAspAsnIlePheGlnIleGlnProMetAsnAsnThrTyTyTyAla 387
Db 949 GAAATCTGAATATGATCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1008
Qy 388 SerGlnIlePheLysSerAsnAspGlyGluAsnIleSerGlyIleCysSerIleGly 407
Db 1009 ACCCAATATTCTTCTAGCAAGCCACCACTGTTTAGGAATACATCAATCGGC 1068
Qy 408 ThrTyArgPheArgLeuGlyGlyAsp---TyrTyArgHisAsnTyLeuValProThr 426
Db 1069 TCATATTCTTTAACTGGGTGATGATGATGATGATGATGATGATGATGATGATGAT 1128
Qy 427 ValLysGlnGlyAsnTyAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheVal 446
Db 1129 ATTAATAATCGAACATTAAGCTTCTATCTGGAATCAACTCTACACATTTGGCTTTGTC 1188
Qy 447 ProValSerGlu 450
Db 1189 CCAGCGCTGAG 1200

```

## RESULT 6

AAZ98631  
ID AAZ98631 standard; DNA; 1440 BP.

AC AAZ98631;

DT 20-JUN-2000 (first entry)

DE Type D BoNT non-toxic immunogenic derivative nucleotide sequence.

DE Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;  
protect; ds.

OS Clostridium botulinum.

PN WO200005252-A1.

XX 03-FEB-2000.

PD

XX 20-JUL-1999; 99WO-IB01301.  
XX 22-JUL-1998; 98ZA-0006538.  
PR (AGRI-) AGRIC RES COUNCIL.  
PA De Bruyn EE, Botha AD;  
XX WPI: 2000-205375/18.  
DR P-PSDB: AAY78982.  
XX  
PT Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin  
PT type D, useful in vaccines for protection against botulism, comprises  
PT at least one amino acid mutation not present in the wild type D  
PT neurotoxins  
XX  
PS Claim 8; Page 59-62; 66pp; English.  
XX  
CC This sequence represents a nucleotide sequence of a synthetic gene  
CC encoding a non-toxic immunogenic derivative of Clostridium botulinum type  
CC D toxin (BoNT). The sequence includes portions of a subtable plasmid which  
CC are immediately upstream and downstream of the gene. Botulinum neurotoxin  
CC causes botulism poisoning in cattle and sheep, and usually results in the  
CC death of the affected or poisoned animal. The non-toxic immunogenic  
CC fragments of the C. botulinum neurotoxin are useful in vaccines to  
CC protect animals (e.g. humans, cattle, sheep, pigs) against BoNT type D  
CC poisoning. The non-toxic fragments can be produced relatively simply and  
CC inexpensively (specifically by fermentation techniques). As the fragments  
CC are not toxic the risk to production staff is reduced.

XX Sequence 1440 BP; 477 A; 262 C; 272 G; 429 T; 0 other;

Alignment Scores:  
Pred. No.: 1.77e-133 Length: 1440  
Score: 1589.00 Matches: 307  
Percent Similarity: 84.16% Conservative: 33  
Best Local Similarity: 75.99% Mismatches: 56  
Query Match: 66.46% Indels: 8  
DB: 21 Gaps: 4

US-09-910-186A-10 (1-450) x AAZ98631 (1-1440)

Qy 50 AlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePropheAspPheLys 69  
Db 58 GCAGAGTTAGAGTTAGGGGAAACGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 117  
Qy 70 LeuGlySerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleVal 89  
Db 118 TTGGGTAGCTCTGGTGATGACCGTGGCAAAATATTGTAACCAACGAAAAATATTGTT 177  
Qy 90 TyrAsnSerMetTyTyGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpVal 109  
Db 178 TATAAGCTATGATGAGTCTTCTCTATTAGCTTCGGATCAGAAATTAATAATGGGTT 237  
Qy 110 SerAsnLeuProGlyTyThrIleIleAspSerValLysAsnAsnSerGlyTrpSerIle 129  
Db 238 TCAAAATTTACCAGGTTATACCATCATCGACAGCGTTAAAAATAATTACGATGGTCTATT 297  
Qy 130 GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer 149  
Db 298 GGCATAATAGCAATTTCTAGTTTTTACACTAAAGCAAAACGAAATACCGACGAGGAC 357  
Qy 150 IleAsnPheSerTyAspIleSerAsnAsnAlaProGlyTyTrpAsnLysTrpPhePheVal 169  
Db 358 ATTAATTTTCTTATGATCATTTCAAGACGCTGCTGTTTATAATAATGTTCTTTGTA 417  
Qy 170 ThrValThrAsnAsnMetMetGlyAsnMetLysIleTyIleAsnGlyLysLeuIleAsp 189  
Db 418 ACCATTACCAACAAACATGATGGGTAATATGATCATCATTAACGAAACGTAATCGAC 477  
Qy 190 ThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle 209



229	ACCATCTACACCAACGACUUCAGTTTGCTCTCCGGTGAC-----AGATCATC	279
82	ValThrGlnAsnGluAsnIleValTyrrAsnSerMetTyrrGluSerPheSerIleSerPhe	101
280	GTCAACCTGGAACAACAACATCTTGTACTCCGCCATCTACGAGAACAATCTCTGTCTCCTTC	339
102	TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyrrThrIleIleAspSer	120
340	TGGATCAAGATCTCCAAGGACTTGACCACTCCCAACAAGAGTACACCATCATCACTCC	399
121	ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu	140
400	ATCGAGCAGAACTCCGGTTGGAAGTTGTATCCGTAAACGGTAAACATCGAGTGGATCTTG	459
141	LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrrAspIleSerAsnAsnAla	160
460	CAGAGCTCAACCGTAAGTACAAGTCCTTGATCTTCGACTACTCCGAGCTCTGTCCTCCAC	519
161	ProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet	179
520	ACCGTTACACCAACAAGTGGTCTTCGTCACCAATCACCAACATCATCGGTTACATG	579
180	LysIleTyrrIleAsnGlyLysIleAspThrIleLysValLysGluLeuThrGlyIle	199
580	AAGTTGTACATCAACCGTGAGTTGAAGCAGTCCCAAGAAGATCGAGGACCTGGACGAGTC	639
200	AsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr	219
640	AAGCTGGACAGACCATCGTCTTCGGTATCGACGAGAACATCGAC-----	684
220	SerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrrIlePheAlaLysGluLeu	239
685	-----GAGAACCATAGTTGTGGATCCGTGACTTCAACTCTTCTCCAAGGAGCTG	735
240	AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrrThrAsnValValLysAsp	259
736	TCCAACGAGGACATCAACATCGTCTACGAGGTCAGATCTCTGAGAACGTCATCAACGAC	795
260	TyrrTrpGlyAsnAspLeuArgTyrrAsnLysGluTyrrTyrrMetValAsnIleAspTyrrLeu	279
796	TACTGGGTAACCCATGAAGTTCGACACCGAGTACTACATCATCAGGACAACTACATC	855
280	AsnArgTyrrMetTyrrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsn	299
856	GACCGTTATACATCCCCAGAGTCACAGTCCTGGTGGCTGCTCCAGTACCTGACCGTTC	915
300	AspPheAsnGluGlyTyrrLysIleIleLysArgIleArgGlyAsnThrAsnAspThr	319
916	AAGCTGTACACCGGTAAACCTATACCATTCAGTCCGCTCTCGACAGAACCCTTACTCC	975
320	ArgValArgGlyGlyAspIleLeuTyrrPheAspMetThrIleAsnAsnLysAlaTyrrAsn	339
976	CGTATCTCTGAACCGTGACAAATCATCTTCGCATCTGTACAACTCCCGTAAGTACATG	1035
340	LeuPheMetLysAsnGlnThrMetTyrrAla-----AspAsnHisSerThrGluAsp	356
1036	ATCATTCGTGACACCCAGACCATCTACGCCACCCAGGCTGGTGTACTGTCCGAACTGT	1095
357	IleTyrrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln	376
1096	GTCATACGCCCTGAAGCTGCAGTCCAACTCCGCTGCTACCGTATCCGGT---ATCTTCTCC	1152
377	IleGlnProMetAsnAsnThrTyrrTyrrTyrrAlaSerGlnIlePheLysSerAsnPheAsn	396
1153	ATCAAGAACATCTGCTCCAGAACAAAGTACTGCTCCAGATCTTC---TCTCTCTCCGT	1209
397	GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyrrArgPheArgLeuGlyGly	415
1210	--GAGAACCATGTCTGTCGCCGACATCTACAGCCTTGGCGTTTCTCC-----	1257
416	AspTrpTyrrArgHisAsnTyrrLeuValProThrValLysGlnGlyAsnTyrrAlaSerLeu	435
1258	-----TTCAAGACCGCTACACTCTCTCGCGCTCACCC-----AACTACGACCAAG	1305

152 CCATATCGAAGGTCCTCATATGCGTAGCATGCTTTATTAAGATATATTAATGATATA 211  
23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVal 43  
212 TTTCAATAGTATTAATGATTCACAAATTTTGGCTTACAAACACAAACAAATGCTTAGT 271  
43 lAspThrSerGlyTyrAsnAlaGluValSerGluGluGluValGlnLeuAsnProIl 63  
272 GCATACATCAGGATATATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 331  
63 ePheProPheAspPheLysLeuGlySerSerGlyLysAspArgGlyLysValIleValth 83  
332 ATATCAATAGTCTTAAATTAAGTAGTCTAGGAGAT-----AAAATTAATGATAA 382  
83 rGlnAsnGluAsnIleValTyrAsnSerMetGlyLysSerPheSerIleSerPheTrpIl 103  
383 TTTAAATATATATATTTATAGCGCTATTTATGAGAACTCTAGTCTAGTTTGGAT 442  
103 eArgIleAsnLysTrpValSerAsnLeuProGly---TyrThrIleIleAspSerVally 122  
443 TAAGATATCTAAAGATTTAACTAATTTCTCATATGCAATATACAAATTAACAGATAGA 502  
122 sAsnAsnSerGlyTyrPheIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGI 142  
503 ACAAAATCTGGTGGAAATATGTTATAGGATGGCAATATAGAAATGGAATTTACAGA 562  
142 nAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaProGI 162  
563 TGTATATAGAAAGTATAAAGATTTAAATTTTATGATATAGTGAATCATTAAGTCATACAG 622  
162 YTYR---AsnLysTrpPhePheValThrValThrAsnMetMetGlyAsnMetLysII 181  
623 ATATCAAAATATGCTTTTGTGTACTATACCAATTAATAATAGGGGTATATGAAC 682  
181 eTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPh 201  
683 TTATATAATAGGAGATTAAGACAGACAGCAAAATTAAGATTTAGATGAGGTAAAGTT 742  
201 eSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAs 221  
743 AGATAAAACCATAGTATTTGGAATAGATGGAATATAGAT----- 782  
221 pSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGI 241  
783 ----GAGATCAGATGCTTGGATAGAGATTTAATATTTTCTAAGAAATTAAGTAA 838  
241 yLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTr 261  
839 TGAAGATATATATGATATGATGGGCAAAATATTAAGAAATGCTTTAAGATTAATG 898  
261 pGlyAsnAspLeuA-gTyrAsnLysGluTyrTyrMetValAsnIleAspTyrIleAsnAr 281  
899 GGGAAATCTTTGAGTTGTACAGATATATATATATATATATATATATATATATAGATAG 958  
281 gTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAspPh 301  
959 GTATATTGCACCTGGAAGATCTACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1018  
301 eAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgva 321  
1019 ATATCTGGAATCTCTATATTAATATAGTATCTGATAGAAATCCTTATAGTAAAT 1078  
321 lArgGlyGlyAspIleLysPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPh 341  
1079 TTTAAATGGAGATATATAATCTTCATATCTATATATATATAGTAGGAATATATAGATAT 1138  
341 eMetLysAsnGluThrMetTyrAlaAspAsnHis-----SerThrGluAspIleTy 358  
1139 AAGAGATCTATACATATATATGCAACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1198  
358 rAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGI 378

1199 TGCATTAATAATTACAGAGTAATTTAGGTATTTAGGTATAGGT---ATATTTAGTATAA 1255  
378 nProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGI 398  
1256 AAATATTGTATCTAAAAATAAATAATTTAGTCTAAATTTTC---TCTAGTTTATAGG---GA 1309  
398 uAsnIleSerGlyIleCysSerIle---GlyThrTyrArgPheArgLeuGlyGlyAspTr 417  
1310 AAATACATGCTTACAGAGATATATATAACCTTGGAGATTCT----- 1355  
417 pTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGI 437  
1356 -TTTAAAAATCATACACGCCAGTTGCAGTAAC-----AATATGAAACAAACTATT 1408  
437 uSerThrSerThrHisTrpGlyPheVal 446  
1409 ATCAACTTCATCTTTTGGAAATTTATT 1436  
RESULT 9  
AAZ87215  
ID AAZ87215 standard; DNA; 1400 BP.  
XX  
AC AAZ87215;  
XX  
DT 08-MAY-2000 (first entry)  
XX  
DE DNA encoding synthetic BoNT serotype E (BoNTE) Hc fragment.  
XX  
KW Botulinum neurotoxin; heavy chain; BoNT; serotype E;  
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
KW VEE; botulism; vaccine; diagnosis; drug screening; ds.  
XX  
OS Clostridium botulinum.  
XX  
Synthetic.  
FH Key Location/Qualifiers  
FT CDS 10..1359  
FT /tag= a  
FT /product= "Synthetic botulinum neurotoxin serotype E  
XX (BoNTE) heavy chain C-terminal fragment (Hc)"  
PN WO200002524-A2.  
XX  
XX 20-JAN-2000. 99WO-US15570.  
XX  
XX 09-JUL-1999; 99WO-US15570.  
XX  
XX 10-JUL-1998; 98US-0092416.  
XX 12-MAY-1999; 99US-0133870.  
XX  
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
XX  
XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;  
XX WPI; 2000-160827/14.  
XX P-PSDB; AAY77137.  
XX  
XX Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
XX toxin serotypes A-G, is used for inducing an immune response against  
XX botulinum -  
XX  
XX Disclosure: Page 43-44; 54pp; English.  
XX  
XX The invention relates to novel vaccines that induce a protective immune  
XX response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
XX and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
XX DNA construct comprising a vector, and at least one nucleic acid  
XX fragment comprising a C-terminal heavy chain fragment (HC) from BoNT  
XX serotypes A-G. In preferred embodiments of the invention, the vector is  
XX a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of  
XX this vector results in the production of large amounts of a protein  
XX encoded by a sequence cloned into the replicon. The constructs are used  
XX to produce vaccines against botulism. The proteins can also be used as

CC diagnostic tools for the diagnosis of botulism. The transformed host  
 CC cells can be used to analyse the effectiveness of drugs and agents which  
 CC inhibit toxin expression. The vaccine currently used against botulism is  
 CC dangerous and expensive to produce, and contains formalin, which is very  
 CC painful for the recipient. Also, the vaccine is incomplete, in that only  
 CC 5 of the 7 serotypes are represented in the formulation. The novel  
 CC vaccine of overcomes these problems, as it is easily purified, and  
 CC available in large quantities. It is also expressed in the lymph nodes  
 CC for a better immune response. Sequences A287212-287217 represent  
 CC synthetic DNA sequences encoding Bont HC fragments used in the present  
 CC invention. These were optimised for codon usage for expression in yeast.  
 XX  
 SQ Sequence 1400 BP; 483 A; 298 C; 232 G; 387 T; 0 other;

## Alignment Scores:

Pred. No.: 1.13e-43 Length: 1400  
 Score: 589.50 Matches: 152  
 Percent Similarity: 51.93 Conservative: 90  
 Best Local Similarity: 32.62 Mismatches: 159  
 Query Match: 24.65 Indels: 65  
 Gaps: 21

us-09-910-186a-10 (1-450) x A287215 (1-1400)

2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
 64 AGTATCCCTTTAACTTCTCTTATACAGATGATAAATTAAATTCCCTACTTCAAC 133  
 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
 124 AAGTCTTCAAGAGAATTAAGTCTCTCTCCGTTTAAACATGAGATACAAGATGATAA 183  
 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
 184 TAGCTCGACACTCCGGTACGATCCCAATACACATTAACGCTGACGCTACAGTAC 243  
 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
 244 CCAACTAACAAAACCAATTC-----GGTATCTACAACGACAGCTTACTGAGCTGAAC 297  
 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
 298 ATCTCTCAAAACGACTACATTATCTACGACACAAAGTACAGAACTCTCTATTCTTTC 357  
 102 TrpIleArgIle-----AsnLysTrpValSerAsnLeuProGlyTyrThrIle 117  
 358 TGGTCAGGATTCCTTAACAGACACAAAGATCGTCAACGTTACACAGGATACACTATC 417  
 118 IleAspSerValLys---AsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
 418 ATCAACTGTATGAGAGACAACTCCGGTTGGAAGTCTCTCTTAACCAACAGAGATC 477  
 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 478 ATTTGGACCTTCAGACACACGCGAGGTATTAAACAAAGTTAGCATTCACCTACGTTAAC 537  
 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
 538 GCAACGGTATTCTTGACTACATCAACAAAGTGGATTTTCGTCACATCACTAACAGACAGA 597  
 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195  
 598 TTAGGTGACTCTACGCTTATCATTAACGGTAACTTAATCGCAAAAGTCCATTTTAAAC 657  
 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215  
 658 TTAGGTAAACATTCACGTTTCTGACAACTCTATTATTAAGATCGTTAACTCGAGTTACAC 717  
 216 GlyLeuIleThrSerAspSerAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235  
 718 AGATACATT-----GGCATTAGATACTTCAACATTTC 750  
 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255

Db 751 GACAGGAGTTAGACGACGAGGATTTTAACTTTATACAGCAACGACCACTAACCAAT 810  
 QY 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275  
 Db 811 ATTTTGAAGGACTTCTGGGTAACACTTGTTCAGCAGCAAGGAATACTTACTTATTAAC 870  
 QY 276 IleAspTyrLeuAsnArgTyrMet-----TyrAlaAsnSerArgGlnIleValPheAsn 293  
 Db 871 GGTGTTAAAGCCAAACAACACTTCATTGATAGGAGAAAGGATTCTACTTTAAGCATTAACAAC 930  
 QY 294 ThrArgArgAsn-----AsnAsnAspPheAsnGluGlyTyrLysIleIleIleLys 310  
 Db 931 ATCAGAGCACTATTTCTTTTAGTAAACAGATTATACCTCTGGTATCAAGGTATAGATCCAA 990  
 QY 311 ArgIle---ArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPhe 329  
 Db 991 AGAGTTAAACAACCTCTCTACTAACGATAACCTTCTAGAAAGACGATCAGGCTATATT 1050  
 QY 330 AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla 349  
 Db 1051 AACITTCGCTGCTAGCAAGACT-----CACTTATTC-----CCATTATATGCT 1092  
 QY 350 AspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIle 369  
 Db 1093 GATACCGCTACCAACAAC-----AAGGAGAGACCATCAAGATC 1131  
 QY 370 AsnAsp-----AsnIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAla 387  
 Db 1132 TCCTCTCTGCAACAGATTTTAAACAAGTCGCTGCTTATGAACCTCCGTCGTAACAACACTGT 1191  
 QY 388 SerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGly 407  
 Db 1192 ACCATGAACCTTTAAATAAT---AATGGAATAAT-----ATTGGG 1230  
 QY 408 ThrTyrArgPheArg-----LeuGlyGlyAspTrpTyrArgHisAsnTyrLeu 423  
 Db 1231 TTGTTAGGTTTCAAGCCAGATCTAGTCTGCTAGTACTTGGTAT----- 1275  
 QY 424 ValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHis--- 442  
 Db 1276 -----TATACCCATGAGATCATCACCAACAGCAATGGA 1311  
 QY 443 -----TrpGlyPheVal 446  
 Db 1312 TGTGTTTGGAACTTTATT 1329  
 RESULT 10  
 AAA54488  
 ID AAA54488 standard; DNA; 1400 BP.  
 XX  
 AC AAA54488;  
 XX  
 DT 11-APR-2001 (first entry)  
 XX  
 DE Botulism toxin heavy chain C-terminal coding sequence (serotype B).  
 XX Botulism; toxin; neurotoxin; heavy chain; recombinant expression;  
 KW recombinant vector; antigen; immune response; vaccine; bacterium;  
 KW infection; ds.  
 XX  
 OS Synthetic.  
 OS Clostridium botulinum.  
 XX  
 FH Key Location/Qualifiers  
 CDS 10..1358  
 FT /\*tag= a  
 FT /product= H\_C peptide fragment  
 XX  
 PN WO200067700-A2.  
 XX  
 PD 16-NOV-2000.  
 XX









Db 3286 ATTTTAAAGACTTTGGGGTGATATTATTAATAATGATGATAAACCACATATATGTTAAAT 3345  
 Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer- 287  
 Db 3346 TTATATGATCAATAAATATGTCGATGTAATATGTTAGGTATATAGAGTTATATGTAAT 3405  
 Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAspPheAsn 302  
 Db 3406 CTTAAAGGGCTAGAGGTAGCGTAATGACTACAAACATTTATTTAAATCAAGTTGTAT 3465  
 Qy 303 GluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322  
 Db 3466 AGGGGGCAAAATTTATTAATAAAATATGCTCTCGAAATAAAGATATATTTGTTAGA 3525  
 Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342  
 Db 3526 AATAATGATCGTATATATTAATGATAGTATGTTAAATAAAGATATAGGTTA----- 3579  
 Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362  
 Db 3580 -----GCTACTAATGTCATCACAGCGGGGTAGAAAATACTAAGT 3621  
 Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382  
 Db 3622 GCATTAGAAATACCTGATGATAGGAATCTTAAGTCAAGTAGTATAGTCAAAATAT 3681  
 Qy 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnLysSerGly 402  
 Db 3682 GATCAAGGAATAACAAATAATGCT---AAATGAATTTACAAGATATTAATGGGAATGAT 3738  
 Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrArgHisAsnTyr 422  
 Db 3739 ATAGGCTTTATAGGATTCATCAGTTT-----AATAAT 3771  
 Qy 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440  
 Db 3772 ATAGCTAACTAGTAGCAAGTAATGTTATATAGCAATAAGAAAGATCTAGTAGGACT 3831  
 Qy 441 -----ThrHisTrpGlyPheValProValSerGlu 450  
 Db 3832 TTGGGTTGCTCATGGGAATTTATTCCTGTAGATGAT 3867  
 Qy 15-FEB-2001 (first entry)  
 Db BoNT/A neurotoxin prototoxin nucleotide sequence SEQ ID NO:10.  
 Human; procholecystokinin; CCK A receptor; CCK B receptor;  
 pancreatitis; antiinflammatory; ss.  
 Clostridium botulinum.  
 WO200061192-A2.  
 19-OCT-2000.  
 06-APR-2000; 2000WO-US09142.  
 08-APR-1999; 99US-0288326.  
 (ALLR ) ALLERGAN SALES INC.  
 Steward LE, Sachs G, Aoki KR;  
 WPI; 2000-679416/66.  
 New composition for treating acute pancreatitis, comprises a pancreatic  
 cell surface marker binding element, a translocation element that

transfers polypeptide across vesicular membrane, and a therapeutic  
 element  
 Disclosure; Page 29-32; 50pp; English.  
 The present invention describes a composition (I) for treating acute  
 pancreatitis. (I) comprises a first element containing a binding element  
 that binds to a pancreatic cell surface marker, a second element  
 containing a translocation element that facilitates polypeptide transfer  
 across the vesicular membrane, and a third element containing a  
 therapeutic element that inhibits enzyme secretion in pancreatic cell  
 cytoplasm. Also described is a method for making a therapeutic  
 polypeptide having a binding element selective for cholecystokinin (CCK)  
 receptor by expressing within a host cell a recombinant chimeric  
 polypeptide comprising an extein containing a therapeutic element and a  
 translocational element, and an intein located to the carboxy terminal  
 of extein having a cysteine, serine or threonine at its amino terminus,  
 and contacting the extein with a synthetic peptide comprising a CCK  
 amino acid sequence containing an amidated phenylalanine at a natural  
 C-terminus, and a cysteine, serine or threonine at its N-terminus, and  
 a nucleophilic reagent able to cause cleavage of the intein to form a  
 peptide bond between the extein C-terminus and synthetic peptide  
 N-terminus through the formation of an activated ester or thio ester  
 intermediate. (I) has antiinflammatory activity and prevents accumulation  
 of pancreatic digestive enzymes, and prevents exocytic fusion of vesicles  
 containing secretory enzymes of pancreas. (I) is useful for treating  
 acute pancreatitis. the present sequence encodes the Clostridium  
 botulinum BoNT/A neurotoxin prototoxin which is given in the  
 exemplification of the present invention.  
 Sequence 4835 BP; 1934 A; 517 C; 756 G; 1628 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 9,45e-43 Length: 4835  
 Score: 586.50 Matches: 149  
 Percent Similarity: 48.94% Conservative: 82  
 Best Local Similarity: 31.57% Mismatches: 186  
 Query Match: 24.53% Indels: 55  
 DB: 21 Gaps: 11  
 US-09-910-186A-10 (1-450) x AAC64582 (1-4835)  
 Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuGlu 22  
 Db 2902 ATACCTTTTCAGCTTCCAAATACGTAGATAATCAAGATTTATTTATCTACATTACTGAA 2961  
 Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGluAsnArgLysAsnThrLeu 42  
 Db 2962 TATATTGAAGATATATTATACTTCTATATTGAATTTAGATATGAAGATATCATTTA 3021  
 Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGlyAspValGlnLeuAsnPro 62  
 Db 3022 ATACACTTATCTAGGTATGCATCAAAATAAATATTGGTAGTAAAGTAAATTTTGATCCA 3081  
 Qy 53 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79  
 Db 3082 ATAGATAAAATCAAAATCAATTAATTTAATTTAGAAGTAGT-----AAA 3126  
 Qy 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99  
 Db 3127 ATTGAGGTAAATTTTAAATAATGCTATTGTATATATATAGTATGTGATAAATTTAGTACT 3186  
 Qy 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116  
 Db 3187 AGCTTTTGGATAAGAATTCCTAAGTATTTTAAACAGTATTAAGTCTAAATATGAATATACA 3246  
 Qy 117 IleIleAspSerValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeu 136  
 Db 3247 ATAATAAATGTATGGAATAATATTCAGGTAGGAAGTATCTACTTAATTTATGGTGAATA 3306  
 Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 3307 ATCTGGACTTTACAGGATCTACTCAGGAATAAACAAGAGAGTAGTTTTTAATATACAGTCAA 3366

```
QY 157 SerAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3367 ATGATTATATATATATATATATATATATATATATATATATATATATATATATAG 3426
QY 176 MetGlyAsnMetLysIleThrIleAsnGlyLysLeuIleAsnThrIleLysValLysGlu 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3427 TTAATACCTCTAAATTTATATATATATATATATATATATATATATATATATAT 3486
QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3487 TTAGCTAATATTCATCTACTACTATAATATATATATATATATATATATATATAT 3546
QY 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetIleArgAspPheThrIlePhe 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3547 -----CATAGATATATTGGATAAATATATATATATATATATATATATAT 3582
QY 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnIleThrAsn 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3583 GATAAGGAATTAATGAAAGAAATCAAGATTTATATGATTAATCAATCAATTCAGGT 3642
QY 256 ValValLysAspTrpGlyAsnAspLeuArgTyrAsnGlyLysIleThrValAsn 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3643 ATTTAAAGACCTTTGGGTGATTTTACAAATATGATAAACCATCTATATATGTTAAAT 3702
QY 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3703 TTATATGATCAATAATATGCGATGTAATAAGTAGGTATAGAGGTATATATGAT 3762
QY 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3763 CTTAAAGGGCGCTAGAGTACGGTAAGTACTACAAACATTTATTTAAATCAAGTTGTAT 3822
QY 303 GluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3823 AGGGGCAAAATTTATTAATAAAATATGCTCTCGAAATAAGATATATTTGTAGA 3882
QY 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3883 AATAATGATCGTGTATATATATATGATAGTATAGTAAATAAAGATATAGTTA 3936
QY 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3937 -----GCTACTAATGCATCACAGCGAGGTAGAGAAATATACTAAGT 3978
QY 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3979 GCATTAGAAATACCTGATGTAGGAAATCTAAGTCAAGTAGTAAATGAATCAAAAAAT 4038
QY 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4039 GATCAAGGAATACAAATAATGC---AAATGAATTCACAGATATATATGGGAATGAT 4095
QY 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyr 422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4096 ATAGGCTTTATAGGATTCATCAGTTT-----AATAAT 4128
QY 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4129 ATACTAACTAGTAGCAACTAATTTGGTATATATAGCAAAATAGAAAGATCTAGTAGGACT 4188
QY 441 -----ThrHisTrpGlyPheValProValSerGlu 450
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4189 TTGGGTGCTCATGGGAATTTATTCCTGTAGATGAT 4224
RESULT 14
AAF58862
ID AAF58862 standard; DNA; 4017 BP.
AC
XX
XX
XX
DT 23-MAY-2001 (first entry)
XX
```

```
DE C botulinum BoNT/E neurotoxin cDNA.
XX
XX Activatable neurotoxin; protease cleavage; H chain; L chain;
KW clostridial neurotoxin; ss.
XX
XX Clostridium botulinum.
XX
XX NC2000114570-A1.
XX
XX 01-MAR-2001.
XX
XX 25-AUG-2000; 2000WO-US23427.
XX
XX 25-AUG-1999; 99US-0150710.
XX
XX (ALLR ) ALLERGAN SALES INC.
XX
XX Dolly JO, Li Y, Chan KC;
XX
XX WPI; 2001-218454/22.
XX
XX Novel isolated single-chain polypeptide derived from activatable
PT recombinant clostridial neurotoxin useful as therapeutic agents,
PT transporter molecules and adducts
XX
XX Example 9; Page 45-47; 50pp; English.
XX
XX The present invention describes an isolated single-chain protein
CC comprising a therapeutic element, and a binding and translocation
CC element. When exposed to a protease, the therapeutic element can be
CC cleaved off. Examples of the binding and translocation element include
CC the clostridial neurotoxin H and L chains. This is useful in the
CC treatment of patients inoculated with the pentavalent BoNT vaccine and
CC for delivering the therapeutic benefits of neurotoxins to patients who
CC are immunologically resistant to a given neurotoxin subtype, patients who
CC may have a lower than average concentration of receptors to a given
CC neurotoxin heavy chain binding group, or patients who may have a
CC protease-resistant variant of the membrane or vesicle toxin substrate.
CC The present sequence was used to demonstrate the actions of the protein
CC of the invention.
XX
XX SQ Sequence 4017 BP; 1673 A; 430 C; 565 G; 1349 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 1,56e-42 Length: 4017
Score: 583.00 Matches: 148
Percent Similarity: 50.86% Conservative: 88
Best Local Similarity: 31.90% Mismatches: 166
Query Match: 24.38% Indels: 62
DB: 22 Gaps: 16
```

US-09-910-186A-10 (1-450) x AAF58862 (1-4017)

```
QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeasn 21
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2592 AGTATTCCTTTTAAGCTTCTCTTATACAGATGATAAAATTTTATTCATATTTTAA 2751
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2752 AAATCTCTTTAAGAGAAATTAAGACTAGTCTCAGTTTAAATATGAGATATAAAATGATAA 2811
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluValAspValGlnLeuAsn 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2812 TAGCTAGATCTCTCAGGATGATGATTCAAATATAAAATATTAATATGAGATGATATATAATAT 2871
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2872 CCACTAATAAATAATCAATTT-----GGAATATATATGATAAACTAGTGAAGTTAAT 2925
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2926 ATATCTCAAAATGATCATATATATATATGATAATAAATAAATAATTTTAGTATTAGTTT 2985
```

[illegible]

```
Db 2626 ATAGATTATCAGGATATGGGCAAGGTAGAGGTATATCATGCGTACGCTTAATGAT 2685
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 53 IlePheProPheAspPheLysLeuGlySerSerGlyCluAspArgGlyLysValIleVal 82
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2686 -----AAAAATCAATTAATAATTAACAGTTCAGCA-----AATAGTAAGATTAGAGTC 2733
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 83 ThrGlnAsnGluAsnIleValTyAsnSerMetTyGluSerPheSerIleSerPheTyr 102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2734 ACTCAAAATCAGATATCATATTAATTAAGTGTGTCCTTCATTTAGCGTTAGCTTTGCG 2793
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 103 IleArgIleAsnLysTyr-----ValSerAsnLeu-----ProGlyTyrThr 116
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2794 ATAAGAATACCTAAATATAAGATGATGATACAAATATATATTCATATGAATATACA 2853
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeu 136
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2854 ATATTAATTGTTAGAAATTAATTCGGGTGGAATATCTATTAGGGGTAAATAGGATA 2913
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyAspIle 156
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2914 ATATGGACTTTAATTGATATAAATGGAATGGAACCAACCAATCGGTATTTTGAATATAACATA 2973
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 157 SerAsnAsnAlaProGlyTyr--AsnLysTyrPhePheValThrValThrAsnAsnMet 175
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2974 AGAGAAGATATATCAGATATATAAATAGATAGGTGTTTGTAACTATTACTATAATAT 3030
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 176 MetGlyAsnMetLysIleTyIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3031 TTGAATACGCTAAATTTATATTAATGTTAGCTAGCAATCAATACAGATATTAAGAT 3090
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheCluIleAsnLysIleProAspThr 215
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3091 ATAGAGAAGATTATGCTAATGTTGGTAATATATTTAAATATGAT----- 3135
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 216 GlyLeuIleThrSerAspSerAsnIleAsn---MetThrIleArgAspPheTyIle 234
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3136 -----GGTGATATAGATAGACACACAATTTATTGGATGAATATTCAGTATT 3183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyThr 254
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3184 TTATAACGGAATTAAGTCAATCAATATTAAGAAAGATATAAAATTCATATATAGC 3243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 255 AsnValValLysAspTyrTyrPheLysAspLeuArgTyrAsnLysGluTyTyrMetVal 274
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3244 GAATATTAAAGATTTTGGGGAATCCCTTAATGTACATAAAGATATATATGTTT 3303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 275 AsnIleAspTyrLeuAsnArgTyrMetTyAlaAsnSer-----ArgGlnIle 290
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3304 AATCGGGGAATAAAATTCATATATTAATACTAAAGAAAGATTCACTGTAGGTGAAT 3363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 291 ValPheAsnThrArgAspAsnAsn-----AspPheAsnGlu 303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3364 TTACACGTACCAATATATCAAAATCTAAATATATAATTAATATAGAGATTATATAT 3423
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 304 GlyTyrLysIleIleIleLysArgIleArgGlyAsnThr-----AsnAspThrArg 320
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3424 GGAGAAAAATTTATATAAAGA--AAGTCAAAATCTCAATCTATAATGATGATATA 3480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 321 ValArgGlyGlyAspIleLeuTyPheAsp--MetThrIleAsnAsnLys----- 336
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3481 GTTAAAAAAGAGATATATATCTAGTATTTTAAATTAATCAAGAGTGGAGAGTA 3540
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 337 ---AlaTyAsnLeuPheMetLysAsnGluThrMetTyAlaAspAsnHisSerThrGlu 355
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3541 TATACCTATAAATATTTAAGAAAGAGAA-----GAA 3573
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 356 AspIleTyAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePhe 375
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3574 AAATGTGTTAGCTCCATTAAGTGTCTGATGAGTTTACAAT-----ACTATA 3624
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 376 GlnIleGlnProMetAsnAsnThrTyTyrTyAlaSerGln---IlePheLysSerAsn 394
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 3625 CAATATAAGAAATATCATGAACCAACACATATAGTTGTCAGTGTCTTTTAAAAA--- 3681
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 395 PheAsnGlyCluAsnIleSerGlyIleCysSerIleGlyThrTyArgPheArgLeuGly 414
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3682 ---GATGAAGAAAGTACTGATGAGATAGGATTGATTGGTATTATTCATCGTTTACGAATCT 3738
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 415 Gly-----AspTyrTyArgHisAsn 421
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3739 GGAATTGCTATTGAGAGATATAAGATTATTTTGTATAAGTAAATGGTAC----- 3789
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 422 TyrLeuValProThrValLysGlnGlyAsnTyAlaSerLeuLeuGluSerThrSerThr 441
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3790 -----TTAAAGAGGTAAAGAAAGAACCAACCATATAATTTAAATTTG-----GGATCT 3834
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 442 HisTyrGlyPheValProValSerGlu 450
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3835 AATGGCAGTTTATTCTCTAAAGATGAA 3861
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: November 7, 2002, 17:51:59  
Job time : 279 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 17:46:26 ; Search time 54 Seconds  
(without alignments)  
2555.640 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTIFNFSTYNNLLKDI.....NYASLLESSTHWGFVPVSE 450

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+p2n.model -DEV=xlh  
/cgn2\_1/uspto.spool/US0910186/runat\_04112002\_111618\_20869/app\_query.fasta\_1.647  
-Isued\_Patents\_NA -QNT=fastcap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPTAT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR.SCORE=ptc -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-CODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0910186.ecgn.1.17 @runat\_04112002\_111618\_20869 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6 -FGAEXT=7  
-XGAPOP=10 -XGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	586.5	24.5	3891	1	US-08-480-604A-27
2	586.5	24.5	3891	2	US-08-405-496A-27
3	586.5	24.5	3891	4	US-08-915-136-27
4	558.5	23.4	1402	1	US-08-480-604A-25
5	558.5	23.4	1402	2	US-08-405-496A-25
6	558.5	23.4	1402	4	US-08-915-136-25
7	557.5	23.3	1330	1	US-08-480-604A-22
8	557.5	23.3	1330	2	US-08-405-496A-22
9	557.5	23.3	1330	4	US-08-915-136-22
10	539.5	22.6	3509	4	US-08-255-829-19
11	423.5	17.7	1858	1	US-08-668-361A-6
12	420	17.6	1359	1	US-07-618-312A-1

#### ALIGNMENTS

##### RESULT 1

US-08-480-604A-27

; Sequence 27, Application US/08480604A

; Patent No. 5736139

; GENERAL INFORMATION:

; APPLICANT: KINK, JOHN A.

; APPLICANT: THALLEY, BRUCE S.

; APPLICANT: PADHYE, NISHA V.

; APPLICANT: FIRCA, JOSEPH R.

; APPLICANT: STAFFORD, DOUGLAS C.

; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND

; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MEDLEN & CARROLL, LLP

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/480,604A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/422,711

; FILING DATE: 14-APR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/405,496

; FILING DATE: 16-MAR-1995

; PRIOR APPLICATION DATA:



; APPLICATION NUMBER: US 08/329,154  
 ; FILING DATE: 25-OCT-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/161,907  
 ; FILING DATE: 02-DEC-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/985,321  
 ; FILING DATE: 04-DEC-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/429,791  
 ; FILING DATE: 31-OCT-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: INGOLIA, DIANE E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: OPND-01763  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 27:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3891 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..3888  
 ; ORF: 480-604A-27

Segment Scores:  
 Score: 3.46e-53 Length: 3891  
 E-value: 586.50 Matches: 149  
 Percent Similarity: 48.94% Conservative: 82  
 Best Local Similarity: 31.57% Mismatches: 186  
 Query Match: 24.53% Indels: 55  
 Gaps: 11

US-09-910-186a-10 (1-450) x US-08-480-604A-27 (1-3891)

3 IleProPheAsnIlePheSerThrAsnAsnSerLeuLeuLysAspIleLeuAsnGlu 22  
 2545 ATACCTTTTCAGCTTCCAAATAGGTAGTAATCAAGATATTATCTACATTTACTGAA 2604  
 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGluAsnArgLysAsnThrLeu 42  
 2605 TATATTAGAAATATTATTAATCTCTATCTATTAATCAATTTAAGATATCAAAAGTAACTATTA 2664  
 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluAspValGlnLeuAsnPro 62  
 2665 ATAGACATATCTAGGTATGCATCAAAATAAATATGGTAGTAAAGTAATTTGTATCCA 2724  
 53 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79  
 2725 ATAGATAAAATCAATCAATCAATTTATTTAATTTAAGAAAGTAGT-----AAA 2769  
 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99  
 2770 ATTGAGGTAAATTTAAATAATGCTATTTGTATATATATATGATGATCAAAATTTTACTACT 2829  
 100 SerPheThrIleArgIleAsnLysThrValSerAsnLeuPro-----GlyTyrThr 116  
 2830 AGCTTTTGGATAAGAAATTCCTAGTATTTTAAACAGTATAAGTCTAAATATGAATATACA 2889  
 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleLeuSerAsnPheLeu 136  
 2890 ATATAAATTTGTATGGAAATAATATCAGTAGGAAGTAGTATCACTTATATGTTGTAATA 2949  
 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 2950 ATCTGGAGCTTTACAGGATCTACAGAAATAAAACAAGAGTAGTATTTTAAATACAGTCAA 3009  
 157 SerAsnAsnAlaProGlyTyr---AsnLysTyrPhePheValThrValThrAsnAsnMet 175

Db 3010 ATGATTAATAGA 3069  
 Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195  
 Db 3070 TTAATAATCTCTAAATTTATATAAATGGAAGATTATAGTCAAAACCAATTTCAAAAT 3129  
 Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215  
 Db 3130 TTAGGTAAATATCTAGTCTAATAATAATAATGTTTAAATATAGATGGTGTAGAGATACA 3189  
 Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235  
 Db 3190 -----CATAGATATATTTGGTAAATATTTTAATCTCTTTT 3225  
 Qy 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnThrThrAsn 255  
 Db 3226 GATAAGGAATTAATGAAAGAAATCAAGATTTATATGATGATCAATCAATCAATCAAGT 3285  
 Qy 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275  
 Db 3286 ATTTTAAAGACATTTGGGTGATTTTACAATAATGATAAACCATACTATATGTTAAAT 3345  
 Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287  
 Db 3346 TTATATGATCCAAATAATAATGTCGATGTAATAATAGTAGGTATTAGAGTTATATGATAT 3405  
 Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302  
 Db 3406 CTTAAAGGGCTAGAGGTACCGTAATGACTACAACATTTATTTAAATCAAGTTGTAT 3465  
 Qy 303 GluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322  
 Db 3466 AGGGGACAAATTTTATATAAAAAATATGCTTCGAAATAAAGATAAATATTTTGA 3525  
 Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342  
 Db 3526 AATAATGATCGTGTATATATATATGATAGTAGTTAAATAAAGATATAGGTTA----- 3579  
 Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362  
 Db 3580 -----GCTACTAATGCTATCATCACAGCGAGCGTAGAAAAATACTAAGT 3621  
 Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382  
 Db 3622 GCATTAAGTAATCTGATGAGGAAATCTAAGTCAAGTAGTAGTATGAAGTCAAAAT 3681  
 Qy 383 ThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402  
 Db 3682 GATCAAGGAATAACAATAAATGC---AAATGAATTTACAAGATAAATAATGGATGAT 3738  
 Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyLysAspTyrTyrArgHisAsnTyr 422  
 Db 3739 ATAGCTTTTATAGGATTTTCATCAGTTT-----AATAAT 3771  
 Qy 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440  
 Db 3772 ATACCTAACTAGTAGCAAGTAATTTGTTATATATAGACAATAAGAAAGATCTAGTAGGACT 3831  
 Qy 441 -----ThrHisTrpGlyPheValProValSerGlu 450  
 Db 3832 TTGGTTGCTCATGGGAATTTTCTCTGTAGATGAT 3867

# RESULT 2

US-08-405-496A-27  
 ; Sequence 27, Application US/08405496A  
 ; Patent No. 5915665  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WILLIAMS, JAMES A.  
 ; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
 ; TITLE OF INVENTION: NEUROTOXIN  
 ; NUMBER OF SEQUENCES: 30  
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN & CARROLL, LLP  
 STREET: 220 MONTGOMERY STREET, SUITE 2200  
 CITY: SAN FRANCISCO  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/405,496A  
 FILING DATE: 16-MAR-1995  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/329,154  
 FILING DATE: 25-OCT-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/161,907  
 FILING DATE: 02-DEC-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/985,321  
 FILING DATE: 04-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/429,791  
 FILING DATE: 31-OCT-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: INGOLIA, DIANE E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: OPD-01308  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 27:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3891 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..3888  
 8-08-405-496A-27  
 Alignment Scores:  
 Seq. No.: 3,468-53 Length: 3891  
 586.50 Matches: 149  
 Percent Similarity: 48.94% Conservative: 82  
 Best Local Similarity: 31.57% Mismatches: 186  
 Every Match: 24.53% Indels: 55  
 Gaps: 11

Db 2770 ATTGAGGTAATTTTAAAAAATGCTATGTATATAAATAGTATGCTATGATAAATTTAGTACT 2829  
 Qy 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116  
 Db 2830 AGCTTTGGGATAAGAATTCCTAAGTATTTTAAACAGTAAAGTCTAAATATGAATATACA 2889  
 Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
 Db 2890 ATAATAATTTGTAAGGAAATAATTCAGGATGGAAGATATCCTTAATTAATGGTGAATA 2949  
 Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 2950 ATCTGGACTTTACAGGATACCTCAGGAATAAACAAGAGTAGTATTTTAAATACAGTCAA 3009  
 Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
 Db 3010 ATGATTATATATCAGATTATATAAACAGATGATTTTGTAACTACTCACTAATATATA 3069  
 Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195  
 Db 3070 TTAATAAATCTCTAAATTTATATAAATGGAAGATTATAGATCAAAAACCAATTTCAAT 3129  
 Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215  
 Db 3130 TTAGGTAATATTCATGCTAGTATAATAATGTTTAAATTAGATGTTGTAGATAGATACA 3189  
 Qy 215 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235  
 Db 3190 -----CATGATATATTTGGATAAAATATTTTAATCTTTT 3225  
 Qy 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255  
 Db 3226 GATAAGGAATTAATGAAAGAAATCAAGATTTATATGATAATCAATCAATCAATCAGGT 3285  
 Qy 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTrpMetValAsn 275  
 Db 3286 ATTTTAAAGACCTTTGGGGTGATTATTTACATAATGATAAACCATACTATATGTTAAT 3345  
 Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287  
 Db 3346 TTATATGATCCAAATAAATATGTCGATGTAATAATAGTATTAGAGTTATATGTTAT 3405  
 Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302  
 Db 3406 CTTAAAGGGCTAGAGTAGGTAATGACTACAAACATTTTATTAAATCAAGTTTGTAT 3465  
 Qy 303 GluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322  
 Db 3466 AGGGGACAAAAATTTATATAAAAAATATGCTTCTGGAATAAAGATAATATTTGTAGA 3525  
 Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342  
 Db 3526 AATAATGATCGTGATATATTAAGTAGTAGTTAAATAAAGAATATAGGTTA----- 3579  
 Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362  
 Db 3580 -----GCTACTAATGCATCACAGGCGAGCGGTAGAAAAATACTACTAAGT 3621  
 Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382  
 Db 3622 GCATTAAGAAATACCTGATGAGGAAATCTAAGTCAAGTAGTAGTAAATGAAGTCAAAAAT 3681  
 Qy 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402  
 Db 3682 GATCAAGGAATAACAAATAAATGC---AAATGAATTTACAAGATAAATATGGGAATGAT 3738  
 Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyr 422  
 Db 3739 ATAGCTTTATAGGATTTTCATCAGTTT-----AATAAT 3771  
 Qy 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440





```
Db 570 CAATAACTCCAAAATCTACATCAAGCGCGTCTGATCGACCAAGAACCGATCTCCAATCT 629
Qy 196 uThrGlyLeuAsnPheserLysThrPheGluLeuAsnLysIleProAspThrCl 216
Db 630 CGGTAACATCCAGCTTCTAATACATCATGTTCAACACGGAGCGTGTGCGGACACT-- 687
Qy 216 yLeuIleThrSerAspSerAspAsnIleAsnMetThrPheArgAspPheThrIlePheAl 236
Db 688 -----CACGCGTCACTCTGGATCGAATACTTCAATCTGTGA 725
Qy 236 aLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnIleThrAsnVa 256
Db 726 CAAGAAGTACGAGAAAGAAATCAAGACCTGTACGACACCGATCCAAATCTGGTAT 785
Qy 256 lValLysAspTyrIlePheGluAsnLysArgIleGluIleGluIleGluIleGluIleGluIle 276
Db 786 CCTCAAGAGACTCTCGGCTGACTACCTGCGAGTACGACAAACCGTACTACATCTGTAATCT 845
Qy 276 eAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287
Db 846 GTACGATCCGAAACAAATACCTGACGTCAACAATGTAGTATCCCGGTTACATGTAATCT 905
Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAspPheAsnGln 303
Db 906 GAAAGGTCGCGTCTGTATGACTACCAACATCTACCTCAATCTCCCTGTACCG 965
Qy 303 uGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGln 323
Db 966 TGGTACCAAAATTCATCATCAAGAAATACGCGTCTGTGTAAACAGGACAAATATGTTGCGAA 1025
Qy 323 yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetIle 343
Db 1026 CAATGATCGGTATACATCAATGTGTAGTATAGAACAAAGATACCGTCTG----- 1077
Qy 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363
Db 1078 -----GCTACCAATGCTCTCAGCGTGTGTAGAAAGATCTTGTCTGC 1121
Qy 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383
Db 1122 TCTGGAATCCGAGCTGTGTAATCTGTCTCAGTGTGTGTATGAATGAATCCAAAGCA 1181
Qy 383 rTyrTyrTyrAlaSerGlnIlePheLysSerAsnAsnGlyGluAsnIleSerGlyIle 403
Db 1182 CCAGGGTATCACTAACAAATGC---AAATGCAATCTGCAGGACAAACAAATGTAACGATAT 1238
Qy 403 eCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyrLe 423
Db 1239 CGGTTTCATCGTTTCCACCATCTC-----AACAAATAT 1271
Qy 423 uValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 1272 CGCTAAACTGGTCTTCCAACTGGTACAAATCGTACATCGTACGACGTTCTCTCGCACTCT 1331
Qy 441 ---ThrHisTrpGlyPheValProValSerGlu 450
Db 1332 GGGTGTCTTGGAGTTCATCCCGGTTGATGAC 1365

RESULT 5
US-08-405-496A-25
; Sequence 25, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUTROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
```

```
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1386
US-08-405-496A-25

Alignment Scores:
Pred. No.: 8,44e-51 Length: 1402
Score: 558.50 Matches: 147
Percent Similarity: 48.73% Conservative: 83
Best Local Similarity: 31.14% Mismatches: 186
Query Match: 23.36% Indels: 56
DB: 2 Gaps: 11

US-09-910-186A-10 (1-450) x US-08-405-496A-25 (1-1402)
Qy 4 PropheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTy 23
Db 45 CCATATCGAAGTCGTCATATGCTAGCATGGCTGCTGCTGCTACCTTCACCTGAATA 104
Qy 23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43
Db 105 CATCAAGACATCATCAATACCTCCATCTCCTGACCTGCGCTAGCAATCCAATCCACCTGAT 164
Qy 43 lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIl 63
Db 165 CGACCTGCTCGCTACGCTTCCAAATCAACATCGCTTCTAAAGTTAACTTCGATCCGAT 224
Qy 63 ePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLysVa 80
Db 225 CGACAAGATCATCATGCTGCTTCAATCTGGAATCTTCC-----AAAT 269
Qy 80 lIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSe 100
Db 270 CGAAGTATCTCTGAAGAATGCTATCTGATACACTTATGTACGAAACTTCTCCACCTC 329
Qy 100 rPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThrIl 117
Db 117 rPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThrIl 117
```

Db 330 CTTCTGGATCCGATATCCGGAATATCTCACTCCATCTCTCTGACAAATACACCAT 389  
QY 117 eileaspSerVallyAsnAsnSerGlyTrpSerIleGlyIleSerAsnPhelLeuVa 137  
Db 390 CATCACTGCAATGGAACAAATCTGGTGGTGAATCTCTGAACACGCTGAAATCAT 449  
QY 137 lPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheserTrpAspIleSe 157  
Db 450 CTGGACTCTGACGAGCATCTAGGAATCAACACGCTGGTGTATTCAAATCTCTCAT 509  
QY 157 rAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnMetMe 176  
Db 510 GATCAACATCTCTGACTACATCTGCTGATCTCTGTTACCATCAACCAATCGTCT 569  
QY 176 tGlyAsnMetLysIleThrIleAsnGlyLysLeuIleAspThrIleLysVallyGluLe 196  
Db 570 GAATACCTCAAAATCTACATACGCGCTGTGTGACGAGAACGATCTCCATCT 629  
QY 196 uThrGlyIleAsnPheserLysThrIleThrPheGluIleAsnLysIleProAspThrGl 216  
Db 630 GGTAAACATCCACGCTTCTAATACATCTTCAAACTGGACGCTTCTGTCGACACT-- 687  
QY 216 yLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTrpIlePheAl 236  
Db 688 -----CACCGCTACATCTGTGATCAAACTACTCAATCTGTGTGCA 725  
QY 236 aLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTrpThrAsnVa 256  
Db 726 CAAGAAGCTGACGAAAGAAATCAAGACCTGTACGACACACGCTCAATCTGTGTAT 785  
QY 256 lVallyAspTrpTrpGlyAsnAspLeuArgTrpAsnLysGluTrpTrpMetValAsnIl 276  
Db 786 CCGTGAAGACTTCTGGGTGACTACCTGCTAGTACGACAAACGCTACTACATCTGTAATCT 845  
QY 276 eaSpTrpLeuAsnAlqTrpMetTrpAlaAspSer----- 287  
Db 845 GTACGATCCGACAAATAGCTGTGACGCAACAAATAGTAGTATCCGCGGTACATGTACCT 905  
QY 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsnGl 303  
Db 906 GAAAGGTCCGCTGCTGTGTATGACTTACCAACATCTACCTGAACCTCTCCCTGTACCG 965  
QY 303 uGlyTrpIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 323  
Db 965 TGTGTACCAATCATCATCAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1025  
QY 323 yGlyAspIleLeuTrpPheAspMetThrIleAsnAsnLysAlaTrpAsnLeuPheMetLy 343  
Db 1026 CAATGATCTGTATACATCAATGTGTAGTATGATTAAGAACAAAGATACCGTCTG 1077  
QY 343 sAsnGluThrMetTrpAlaAspAsnHisSerThrGluAspIleTrpAlaIleGlyLeuAr 363  
Db 1078 -----GCTACATGCTTCTCAGGCTGGTGTAGAAAGATCTGTCTGTC 1121  
QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383  
Db 1122 TCTGGAATCCCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1181  
QY 383 rTrpTrpTrpAlaSerGlnIlePheLysSerAspPheAsnGlyLysAsnIleSerGlyIl 403  
Db 1182 CCAGGTATCTACTACAAATG---AAATGAATCTGACGACCAACATGTTACAGGATAT 1238  
QY 403 eCysSerIleGlyThrTrpArgPheArgLeuGlyGlyAspTrpTrpArgHisAsnTrpLe 423  
Db 1239 CGGTTTCATCGGTTCCACGAGTTC-----AACATAT 1271  
QY 423 uValProThrVallyGlnGlyAsnTrpAlaSerLeuLeuGluSerThrSer----- 440  
Db 1272 CCGTAACTGGTGTCTTCCAACTGGTGAATCGTACGATCGACGATCTCTCTCGCACTCT 1331  
QY 441 ----ThrHisTrpPheValProValSerGlu 450  
Db 1332 GGGTGTCTCTGGAGTTCATCCCGGTTGATGAC 1365

## RESULT 6

US-08-915-136-25  
Sequence 25, Application US/08915136  
Patent No. 6290960  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,136  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,604  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPND-01763  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1402 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1386  
US-08-915-136-25  
Alignment Scores:  
Pred. No.: 8.44e-51 Length: 1402  
Score: 558.50 Matches: 147  
Percent Similarity: 48.73% Conservative: 83  
Best Local Similarity: 31.14% Mismatches: 186  
Query Match: 23.36% Indels: 56

DB: 4 Gaps: 11  
US-09-910-186A-10 (1-450) x US-08-915-136-25 (1-1402)  
Qy 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLeuAspIleleAsnGluTyr 23  
Db 45 CCATATCGAAGGTGCTATATGCTAGCATGGCTGCTGCTGCTACCTTCACTGAAT 104  
Qy 23 rPheAsnIleAsnAspSerTyrIleLeuSerLeuGlnAsnArgIlyAsnThrLeuVa 43  
Db 105 CATCAAGAACATCATCAATACCTCCATCCCTGAACTGCGCTAGCAATCAACCATCAT 164  
Qy 43 LaspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIle 63  
Db 165 CGACTGCTCGCTACCTGCTCCAAATACATCGGTTCTAAAGTTAACTTCGATCCGAT 224  
63 ePheProPheAsp-----PheIysLeuGlySerSerGlyGluAspArgGlyIysVa 80  
225 CGACAAGAAATCAGATCCAGCTGTTCAATCTGGAATCTTCC-----AAAT 269  
80 IrlleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSe 100  
270 CGAAGTTATCTGGAAGATGCTATCGTATCAACTCTATGACGAATACTTCCACCTC 329  
100 rPheTrpIleArgIleAsnIlyrPvalSerAsnLeuPro-----GlyTyrThrIle 117  
330 CTCTGATCGTATCCGAAATCTTCAACTCCATCTCTGAAACAATGAATACACCAT 389  
117 eIleAspSerValIysAsnAsnSerGlyTyrPheIleGlyIleIleSerAsnPheLeuVa 137  
390 CATCACTGCTGGAAGAAATCTGTTGGAAGATCTCTGAACTACGAGTGAATCAT 449  
137 lPheThrLeuIysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSe 157  
450 CTGGACTCTGAGGACACTCAGGAATCAACACGCGTGTATTCAATACTCTCAGAT 509  
157 rAsnAsnAlaProGlyTyr---AsnIlyrPhePheValThrValThrAsnAsnMetMe 176  
510 GATCAACATCTCTGACTACATCATCTGCTGGATCTCGTTACCATCCCAACATGCTCT 569  
176 tGlyAsnMetIlyrIleAsnGlyIysLeuIleAspThrIleIysValIysGluLe 196  
570 GAATAACCTCCAAATCTACATCAACGCGCTGATCGACCAAGAACCGATCTCCAATCT 629  
196 uThrGlyIleAsnPheSerIlyrThrIlePheGluIleAsnIlyleProAspThrGl 216  
630 GGGTAACATCCACCTCTTAATACATCATCTGCTCAACTGAGCGGTGCTGACACT-- 687  
216 yLeuIleThrSerAspSerAsnIleAsnMetTyrIleArgPheTyrIlePheAl 236  
688 -----CACCGCTACATCTGGATCAATATCTTCAATCTGTTGCA 725  
236 aIysGluLeuAspGlyIysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnVa 256  
726 CAAGAAGCTGAACGAAAGAAATCAAGACCTGTACGACACACCGATCCCAATCTGGTAT 785  
256 lValIysAspTyrTrpGlyAsnAspLeuArgTyrAsnIlyGlyTyrMetValAsnIle 276  
786 CCGAAGACTCTCGGGTGACTACCTTCGAGTACGAAACCGTACTACTGCTGATCT 845  
276 eAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287  
846 GTACGATCCGAAACAAATACGTTGAGCTCAACAATGTAGGTATCCGCGGTATACATGTACCT 905  
288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsnGl 303  
906 GAAAGGTCCGGTGGTGTGTTGATGACTACCAACATCTACTGAACTCTCCCTGTACCG 965  
303 uGlyTyrIlyleIleIleIysArgIleArgGlyAsnThrAsnAspThrArgValArgGl 323  
966 TGGTACCAAAATTCATCATCAAGAAATACGGTCTGCTACAGCAACATATCGTTCCAA 1025  
323 yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnIlyAsnAlaTyrAsnLeuPheMetIly 343

Db 1026 CAATGATCGGTATACATCAATGTTGTAGTTAAGAACAAAGATACCGTCTG----- 1077  
Qy 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363  
Db 1078 -----GCTACCAATGCTTCTCAGGCTGGTGTAGAAAGATCTTGTCTGC 1121  
Qy 363 gGluGlnThrIlyAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383  
Db 1122 TCTGGAATCCCGAGCTGGTGTATCTGTCTCAGGTAGTGTAAATGAATCAAGAAGCA 1181  
Qy 383 rTyrTyrTyrAlaSerGlnIlePheIysSerAsnPheAsnGlyGluAsnIleSerGlyIle 403  
Db 1182 CAGGGTATCACAACAATGC---AAATGATCTGACGACACAAATGGTAAGATAT 1238  
Qy 403 eCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsnTyrLe 423  
Db 1239 CGGTTTCATCGGTTTCCACCATCTC-----AACAAATAT 1271  
Qy 423 uValProThrValIysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440  
Db 1272 CGCTAACTGTTGCTTCCAACTGGTACATCGTCAGATCGAAGTCTCTCTCGCACTCT 1331  
Qy 441 ---ThrHisTrpGlyPheValProValSerGlu 450  
Db 1332 CGGTTGCTCTGGAGTTCATCCCGTTGATGAC 1365

RESULT 7  
US-08-480-604A-22  
; Sequence 22, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIORITY DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIORITY DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIORITY DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIORITY DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIORITY DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIORITY DATA: US 07/429,791

FILING DATE: 16-MAR-1995  
REFLECTION NUMBER: 95/3

Copied from 10910186 on 05-05-2004



Copied from 10910135 on 05-05-2004

NUMBER OF SEQUENCES: 32  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MEDLEN & CARROLL, LLP  
 STREET: 220 MONTGOMERY STREET, SUITE 2200  
 CITY: SAN FRANCISCO  
 STATE: CALIFORNIA  
 COUNTRY: UNITED STATES OF AMERICA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/915,136  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/480,504  
 FILING DATE:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/405,496  
 FILING DATE: 16-MAR-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/329,154  
 FILING DATE: 25-OCT-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/161,907  
 FILING DATE: 02-DEC-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/985,321  
 FILING DATE: 04-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/429,791  
 FILING DATE: 31-OCT-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: INGOLIA, DIANE E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: OPND-01763  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 22:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1330 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1314  
 US-09-915-136-22

Copied from 10910186 on 05-05-2004

Alignment Scores:  
 Pred. No.: 1e-50 Length: 1330  
 Score: 557.50 Matches: 144  
 Percent Similarity: 48.80% Conservative: 80  
 Best Local Similarity: 31.37% Mismatches: 180  
 Query Match: 23.32% Indels: 55  
 DB: 4 Gaps: 11  
 US-09-910-186a-10 (1-450) x US-08-915-136-22 (1-1330)  
 QY 16 LeuLysAspIleAlaGlyLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlu 35  
 DB 10 CTGCTGTCTACCTTCCAGTGAATCATCAAGAACATCATCAATACCTCCCTGAACCTG 69  
 QY 36 GlnAsnArgLysAsnThrValAspThrSerGlyTyrAsnAlaGluValSerGluGlu 55  
 DB 70 CGCTAGCAATCAATCACTGATGACCTGTCTCGTACGCTTCCAAATCAACATCGGT 129  
 QY 56 GlyAspValGlnLeuAsnProIlePheProPheAsp-----PheLysLeuGlySer 72

DB 130 TCTAAAGTTAACTTCGATCGCATCGACAGCAAGAAATCAGATCCAGCTGTTCATCTGGAAATC 189  
 QY 73 SerGlyGluAspArgGlyLysValIleValThrGluAsnGluAsnIleValTyrAsnSer 92  
 DB 190 TCC-----AAATCGAAGTTATCCTGAAGATCTATCGTATACACTCT 234  
 QY 93 MetTyrGluSerPheSerIleSerPheTyrIleArgIleAsnLysTrpValSerAsnLeu 112  
 DB 235 ATGTACGAAATCTCTCCAGCTCTCTGATCGGTATCCGAAATACTTCACTCCATCCATC 294  
 QY 113 Pro-----GlyTyrThrIleLeuAspSerValLysAsnAsnSerGlyTyrSerIle 129  
 DB 295 TCTCTGAACAATGAATACACATCATCACTGCATGGAACAAATCTGGTGGGAAGTA 354  
 QY 130 GlyIleLeuSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer 149  
 DB 355 TCTCTGAACACGCTGGAATCATCTGGACTCTCGAGACACTCAGGAATCAACACAGCT 414  
 QY 150 IleAsnPheSerTyrAspIleSerAsnAlaProGlyTyr--AsnLysTrpPhePhe 168  
 DB 415 GTTGTATTCAAAATCTCTCAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTTC 474  
 QY 169 ValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIle 188  
 DB 475 GTTACCATCACCAACAATCGTCTGAATACTCCAAATCTACATCAACGGCGCTGTGATC 534  
 QY 189 AspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGlu 208  
 DB 535 GACCAGAAACGATCTCCAATCTGGGTAACTCCAGCTTCTTAATACATCATGTGTCAA 594  
 QY 209 IleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrp 228  
 DB 595 CTGGACGCTTCTCGTACACT-----CACCGCTACATCTGG 630  
 QY 229 IleArgaspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe 248  
 DB 631 ATCAAACTTCAATCTGTTCGACAAAGAACTGACGAAAGAAATCAAGACCTGTGATC 590  
 QY 249 AsnSerLeuGlnTyrThrAsnValLysValLysAspTyrTrpGlyAsnAspLeuArgTyrAsn 268  
 DB 691 CACAACAGTCCTCAATCTCGTATCTCTGAAAGACTCTCTGGGTGACTACCTGACGACAG 750  
 QY 269 LysGluTyrTyrMetValAsnIleAspTyrIleAsnArgTyrMetTyrAlaAsnSer--- 287  
 DB 751 AAACCGTACTACATGCTGAATCTGTACGATCCGAAACAATACGTTGACGTCAACAATGA 810  
 QY 288 -----ArgGlnIleValPheAsnThrArg--- 295  
 DB 811 GGTATCCCGGTTACATGTACCTGAAAGTCCCGCTGTTCTGTATGACTACCAACATC 870  
 QY 296 ArgAsnAsnAsnAspPheAsnGluGlyTyrIleIleIleLysArgIleArgGlyAsn 315  
 DB 871 TACCTGAACCTCTCCCTGTACCGGTGACCAAAATTCATCATCAAGAAATACGCGCTGTGT 930  
 QY 316 ThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsn 335  
 DB 931 AACAGGACAATATCGTTCGCAACAATGATCGGTATACATCAATGTTCTAGTAGAAGAC 990  
 QY 336 LysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGlu 355  
 DB 991 AAAGAATAACCGCTGTG-----GCTACCAATGCTTCTCAGCT 1026  
 QY 356 AspIleTyrAlaIleGlyLeuArgGluThrLysAspIleAsnAspAsnIleIlePhe 375  
 DB 1027 GGTGTAGAAACATCTGTCTGCTCGTCTGGAATCCCGGACGCTGTGTAATCTGTCTCAGGTA 1086  
 QY 376 GlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPhe 395  
 DB 1087 GTTGTGAATGAATCAACACACGACGAGGTATCACTAACAAATGC---AAATGAATCTG 1143  
 QY 396 AsnGlyGluAsnIleSerGlyIleCysIleGlyThrTyrArgPheArgLeuGlyGly 415

Db 1144 CAGGACAAATGTAACGATATCGGTTTCATCGGTTTCCACCGTTC----- 1191  
Qy 416 AspTrpYrargHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu 435  
Db 1192 -----AACAAATATCGTAACACTGGTTGCTTCCAACTGGGACAAATCGTCAG 1236  
Qy 436 LeuGluSerThrSer-----ThrHisTrpGlyPheValProValSerGlu 450  
Db 1237 ATCGAACGTTCCCTCGCACTCTCGGTTGCTCTTGGAGTTTCATCCGGTTTGATGAC 1293

## RESULT 10

US-09-255-829-19  
; Sequence 19, Application US/09255829  
; Patent No. 6461617  
; GENERAL INFORMATION:  
; APPLICANT: Shone, Clifford Charles  
; APPLICANT: Quinn, Conrad Padraig  
; APPLICANT: Foster, Keith Alan  
; APPLICANT: Rebinbant Toxin Fragments  
TITLE OF INVENTION: Recombinant Toxin Fragments  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN, & FOX P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20005-3934

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/255,829  
FILING DATE: 23-FEB-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB97/02273

FILING DATE: 22-AUG-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/782,893

FILING DATE: 27-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: ESMOND, ROBERT W.

REGISTRATION NUMBER: 32,893

REFERENCE/DOCKET NUMBER: 1581.0130002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 3509 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..3509

US-09-255-829-19

Alignment Scores:

Pred. No.: 3,38e-48 Length: 3509  
Score: 539.50 Matches: 130  
Percent Similarity: 55.75% Conservative: 64  
Best Local Similarity: 37.35% Mismatches: 117  
Query Match: 12 Indels: 37  
DB: 4 Gaps: 12

US-09-910-186a-10 (1-450) x US-09-255-829-19 (1-3509)

Qy 3 ileProPheAsnIlePheSerTyrThrAsnSerLeuLeuLysAspIleAsnGlu 22

Db 2506 ATGCGTTTGATCTTCAATATATACCAATGATACAACTCTAATAGAAATGTTAATAAA 2565

Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42  
Db 2566 TATAATAGCGAAATTTAATATATATCTTAATTTAAGTATATAGGATATAGGATATATTA 2625  
Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62  
Db 2626 ATAGATTATCAGGATATGGCGCAAGAGTAGAGGTATATATGATGGAGTCGAGCTTAATGAT 2685  
Qy 63 IlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleVal 82  
Db 2686 -----AAAAATCAATTTAAATTAAGTATAGCTCAGCA-----AATAGTAAGATTAGAGTG 2733  
Qy 83 ThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrp 102  
Db 2734 ACTCAAAATCAGAAATATCATATTTAATAGTGTGCTTTCCTTGAATTTAGCGTTAGCTTTGG 2793  
Qy 103 IleArgIleAsnLysTrp-----ValSerAsnLeu-----ProGlyTyrThr 116  
Db 2794 ATAAGAATACCTAAATATAGAAATGATGATATACAAATATATATTCATATAGAAATATACA 2853  
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
Db 2854 ATAATTAAATTGTATGAAATAATTCGGGCTGCGGCTGAAATATATCTATTAGGGTAATAGGATA 2913  
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerIleAspIle 156  
Db 2914 ATATGGACTTTAATTTGATATATAAATGGAACCAATCGGTATTTTGTGAATATAACATA 2973  
Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
Db 2974 AGAAGAATATATCAGATATATAAATAGATGCTTTTGTGAACCTATTACTAATAAT--- 3030  
Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195  
Db 3031 TTGATAACCGCTAAATTTATATTAATGTAAGCTAGATCAATCAATACAGATATTAAAGAT 3090  
Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrGluIleAsnLysIleProAspThr 215  
Db 3091 ATAAGAGAAGTTATTGCTAATGCTGAATATAATATTAAATTAGAT----- 3135  
Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsn---MetTrpIleArgAspPheTyrIle 234  
Db 3136 -----GGTGATATAGATAGACACAACTTTATTTGGATGAAATATTTTCAGTATT 3183  
Qy 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThr 254  
Db 3184 TTTAATACGGATTAAGTCAATCAATATTTGAAGAAGATATAAATCAATCATATAGC 3243  
Qy 255 AsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetVal 274  
Db 3244 GAATATTAAAGATTTTGGGGAATCCCTTTAATGTACATAAAGAATATATATATGTTT 3303  
Qy 275 AsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer-----ArgGlnIle 290  
Db 3304 ATCGCGGGGAATAAATTCATATATTAATACTAAAGAAGATTCACCTGTAGGTGAATTT 3363  
Qy 291 ValPheAsnThrArgArgAsnAsnAsn-----AspPheAsnGlu 303  
Db 3364 TTAACACGTAGCAATATAATCAAAATTTCTAAATATATAAATATAGAGATTATATATT 3423  
Qy 304 GlyTyrLysIleIleIleLysArgIleArgGlyAsnThr-----AsnAspThrArg 320  
Db 3424 GGAGAAAAATTTATTATAGAGA---AAGTCAAAATTCCTCAATCTATAAATGATGATATA 3480  
Qy 321 ValArgGlyGlyAspIleLeuTyr 328  
Db 3481 GTTAGAAAAAGAGATATATATATAT 3504

## RESULT 11

US-08-668-381A-6  
; Sequence 6, Application US/08668381A  
; Patent No. 5780024



Qy 416 Asp-----TptYrAIGHisAsnTyrlLeuValPro 425  
 Db 1742 GATCCAAATAGGATATATTAATGCAACACTGGTAC---TTTAAATCAATTTAAAGAT 1798  
 Qy 426 ThrValysGlnGlyAsnTyrlAlaSerLeuLeuGluSerThrHisTrpGlyPhe 445  
 Db 1799 AAATTT-----TTAGGATGTGATTTGGTACTTT 1825  
 Qy 446 ValProValSerGlu 450  
 Db 1826 GTACCTACAGATGAA 1840

RESULT 12

US-07-618-312A-1  
 ; Sequence 1, Application US/07618312A  
 ; Patent No. 5389540

GENERAL INFORMATION:

APPLICANT: Makoff Dr., Andrew J  
 APPLICANT: Romanos Dr., Michael A  
 APPLICANT: Claire Dr., Jeffrey J  
 APPLICANT: Fairweather Dr., Neil F  
 TITLE OF INVENTION: VACCINES  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: 14th Floor  
 STREET: 2200 Clarendon Boulevard,  
 CITY: Arlington,  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/618,312A  
 FILING DATE: 19910516  
 CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 8926832.0  
 FILING DATE: 28-NOV-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: GB 9006097.1  
 FILING DATE: 17-MAR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Crawford Mr. Arthur R  
 REGISTRATION NUMBER: 25,327  
 REFERENCE/DOCKET NUMBER: 510-51  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 0101 703 8750400  
 TELEFAX: 0101 703 5253468  
 TELEX: 200797 NIXN UR

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1359 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Clostridium tetani  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1356  
 US-07-618-312A-1

Alignment Scores:

Pred. No.: 6.26e-36 Length: 1359  
 Score: 420.00 Matches: 131  
 Percent Similarity: 44.20% Conservative: 90  
 Best Local Similarity: 26.20% Mismatches: 169

Query Match: 17.57% Indels: 110  
 DB: 1 Gaps: 20  
 US-09-910-186A-10 (1-450) x US-07-618-312A-1 (1-1359)  
 Qy 6 AsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsn 25  
 Db 7 AATCTGGATTGTGGTTCATATGAAGAAGATATAGATTATATA----- 51  
 Qy 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45  
 Db 52 ---TTAAAAAGAGTACAATTTTAAATTAGATATATAATAGATATTAATATCGATATA 108  
 Qy 46 SerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePro 65  
 Db 109 TCTGGGTTTAATTCATCTGTAATAACATATCCAGATGCTCAATTGGTGCCC----- 159  
 Qy 66 PheAspPheLysLeuGlySerSerGlyGlu-----AspArgGly 78  
 Db 160 -----GGAATAAATGGCAAGCAACATACATTTAGTAAACAATGAATCTTCT 204  
 Qy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98  
 Db 205 GAAGTTATATGTCATAAAGCTATGATATTGAATATATATGATATGTTTAAATAATTTACC 264  
 Qy 99 IleSerPheTrpIleArgIleAsnLys---TtpValSerAsnLeuPro----- 113  
 Db 265 GTTAGCTTTTGGTGGAGGTTCTCTAAAGTATCTGCTAGTCATTAGAACATATAGGCACA 324  
 Qy 114 ---GlyTyrThrIleIleAspSerValLysAsnAsn-----SerGlyTrp 127  
 Db 325 AATGAGTATTCAATAATTAGCTCTATGAAAAAACATAGTCTATCAATAGGATCTGGTTGG 384  
 Qy 128 SerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGlu 147  
 Db 385 AGTGATCACTTAAAGTAAATAACTTAATATGAGCTTTAAAGATTCGCCGGGAGAAATT 444  
 Qy 148 GlnSerIleAsnPheSerTyrAspIleSerAsnAlaProGlyTyr-----AsnLys 165  
 Db 445 AGACAATAACTTTTAGG---GATTTACCTGATAATTTAATGCTTATTAGCAATAAA 501  
 Qy 166 TrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly 185  
 Db 502 TGGTTTTTATACTATTACTATATGATAGTATCTCTGCTAAATTCGTATATATAATGA 561  
 Qy 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205  
 Db 562 GTACTTATGGGAAGTCAGAGAAATTAGTGGTTAGGAGCTATTAGAGAGGATAATAATA 621  
 Qy 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIle 225  
 Db 622 ACATTAATAACTAGATAGA-----TGTAATAATAATAATAACA 657  
 Qy 226 AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspLysLysAspIleAsn 245  
 Db 658 TAGCTTCTATTGATAAATTTAGGATATTTCCAAGCATTAATCCCAAGAGATTGAA 717  
 Qy 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAsnAspIle 265  
 Db 718 AAATTATACACAAGTTATTATCTATAACCTTTTAAAGAGACTTCTGGGAAACCCCTTAA 777  
 Qy 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276  
 Db 778 CGATATGATACAGAATATTATTAAACAGCAGCTTCCTAGTTCCTAAAGATCTTCANTTG 837  
 Qy 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289  
 Db 838 AAAAAATATACAGATTATATGTTTGGCAAAATGGCCCATCGTATACTAACGCAAAATTTG 897  
 Qy 290 IleValPheAsnThrArgAsnAsnAsnAspPheAsnGluGlyTyrIleIleIle 309  
 Db 898 AATATATATTATAGAAGGTTATATAAT-----GGACTAAATATTATTATA 942

QY 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyGlyAspIleLeuTyr 328  
Db 943 AAAAGATACACCTATATATGAATAGTCTTTGTTAAATCAGGTGATTTTATAA 1002  
QY 329 PheAspMetThrIleAsnAsnLysAlaTyr-----AsnLeu 340  
Db 1003 TTATATGATCATATAACAATAATAGCACATCTAGGTTATCCGAAAGATGGAATGCC 1062  
QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
Db 1063 TTTAATATCTGTAAGATCTTAAGAGTAGTATTAATGCCCCAGGTATCCCTCTTTAT 1122  
QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPromet 380  
Db 1123 AAAAAATGGAAGCAGTAAAA-----TTG 1146  
381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400  
1147 CGTGATTAAACCACTATTCTGACAACTT-----AAATTATATGATGATAAAATGCA 1200  
401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAsp-----416  
1201 TCT---TTAGGACTAGTAGTACCCATAAGTCAATAGGCAACGATCCCAATAGGGAT 1257  
417 -----TptTyrArgHisAsnTyrLeuValProThrValLysGlnGly 430  
1258 ATATTAATGCAACCACTGGTAC---TTTATATCAATTTAAAGATAAAATT-----1305  
431 AsnTyrAlaSerLeuLeuGluSerThrSerThrHisTyrGlyPheValProValSerGlu 450  
1306 -----TTAGGATGTCATGGTACTTGTACCTACACATGAA 1341

Copyright 1994-2004

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1359 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Clostridium tetani  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1356  
US-07-618-312A-3  
Alignment Scores:  
Pred. No.: 6,266-36 Length: 1359  
Score: 420.00 Matches: 131  
Percent Similarity: 44.20% Conservative: 90  
Best Local Similarity: 26.20% Mismatches: 169  
Query Match: 17.57% Indels: 110  
Gaps: 20  
US-09-910-186A-10 (1-450) x US-07-618-312A-3 (1-1359)  
QY 6 AsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsn 25  
Db 7 AACCTTCATGTTGGTTCGACACGAAGAGACATCGATGTTATC-----51  
QY 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45  
Db 52 ---CTGAAAGTCTACCATCTGAACTGGACATCAACAACGATATATATCCGACATC 108  
QY 46 SerGlyTyrAsnAlaGluValSerGluGluValGlyAspValGlnLeuAsnProIlePhePro 65  
Db 109 TCUGTTTCACCTCCCTGTTATCATATATCCAGATGCTCAATTTGGTCCG-----159  
QY 66 PheAspPheLysLeuGlySerSerGlyGlu-----AspArgGly 78  
Db 160 -----GGCATCAACGCGCAAGCTATCCACTGGTTAAACAACGAATCTTCT 204  
QY 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98  
Db 205 GAATTTATCTGCACAGGCCATGGACATCGAATACAGACATGTTCAACACTTCACC 264  
QY 99 IleSerPheTrpIleArgIleAsnLys---TrpValSerAsnLeuPro-----113  
Db 265 GTTAGCTTCGGCTGCGCTCCGAAAGTTTCTGCTCCCACTGGGAACAGTACGCGCACT 324  
QY 114 ---GlyTyrThrIleAspSerValLysAsnAso-----SerGlyTrp 127  
Db 325 AACGAGTACTCCATCATCAGCTCTATGAAGAAACACTCCCTGTCACCTGGCTCTGTTGG 384  
QY 128 SerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGlu 147  
Db 385 TCTGTTCCCTGAAGGGTAAACAACCTGATCTGGACTCTGAAAGACTCCGCGGCGAAGTT 444  
QY 148 GlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr-----AsnLys 165  
Db 445 CGTCAGATCACTTTCCGC---GACCTGCGCGCAAGATTCAACGCGTACCTGGCTAACAAA 501  
QY 166 TrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly 185  
Db 502 TGGGTTTTCATCACTATCACTAAGCATCGTCTGCTCTCTCTAACTGTACATCAACGCG 561  
QY 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205  
Db 562 GTTCTGATGGCTCCGCTGAAATCACTGGTCTGGCGCATCCGTGAGGACAAACAACATC 621  
QY 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIle 225  
Db 622 ACTCTTAAGCTGGACCGT-----TGCAACAACAACAACAG 657

```
Oy 226 AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspClyLysAspIleAsn 245
Db 658 TAGTCTCCATCGACAGAGTTCCGTACTCTTCGCAAGACACTGAACCCGAAAGAGATCGAA 717
Oy 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpClyAsnAspLeu 265
Db 718 AAACGTATACCACTACCTGTCTATACCTCTCTCGGTGACTTCTGGGTAAACCCGCTG 777
Oy 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276
Db 778 CGTTACGACCGCAATATACCTGACCGGTAGCTTCTAGTCTTAAGACGTTCCAGGTG 837
Oy 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289
Db 838 AAAAATCATCTACCTACCTACCTGACCAACGCGCGCTCTACCTAAACGGTAACTG 897
Oy 290 IleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyrIleIleIle 309
Db 898 AACATCTACCTACCGACGCTGTGACAAC-----GGCCTGAATTCATCATC 942
Oy 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyLysAspIleLeuTyr 328
Db 943 AAACGCTACACTCCGAACAACGAATCTCTTCGTTAAATCTGGTGACTCATCAAA 1002
Oy 329 PheAspMetThrIleAsnAsnLysAlaTyr-----AsnLeu 340
Db 1003 CTGTACGTTTCTTACACAACAACGACATCGTTGGTTACCCGAAAGACGGTAACGCT 1062
Oy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
Db 1063 TTCAACAACCTGGACAGAACTCTGGGTGTGTACACGCTCCGGGTATCCCGTGTAC 1122
Oy 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIlePheGlnIleGlnProMet 380
Db 1123 AAAAAAATGAAGCTGTAAA-----CTG 1146
Oy 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Db 1147 CGTGACCTGAACCTACTCTGTACGTG-----AACTGTACGACGACAAACGCT 1200
Oy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyLysP----- 416
Db 1201 TCT---CTGGGTCTGTGTGTACCAACGCTGACATCGTAAACGCGGACCGGACG 1257
Oy 417 -----TrpTyrArgHisAsnTyrLeuValProThrValLysGlnGly 430
Db 1258 ATCTGTACGCTCTACTACGTGTAC---TTCACCCACCTGAAGACAAATC----- 1305
Oy 431 AsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheValProValSerGlu 450
Db 1306 -----CTGGGTGCGACTGTCTACTTCTCGGACCGATGAA 1341
```

RESULT 14

US-08-280-228-1

Sequence 1, Application US/08280228

Patent No. 5571694

GENERAL INFORMATION:

APPLICANT: Makoff Dr, Andrew J

APPLICANT: Romans Dr, Michael A

APPLICANT: Clare Dr, Jeffrey J

APPLICANT: Fairweather Dr, Neil F

TITLE OF INVENTION: VACCINES

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON &amp; VANDERHYE P.C.

STREET: 1100 No. 5571694th Glebe Road

CITY: Arlington,

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/280,228
FILING DATE: 25-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/618,312
FILING DATE: 27-NOV-1990
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8926832.0
FILING DATE: 28-NOV-1989
APPLICATION DATA:
APPLICATION NUMBER: GB 9006097.1
FILING DATE: 17-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mary J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 117-163
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
TELEX: 200797 NIXN GR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1359 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORGANISM: Clostridium tetani
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1356
US-08-280-228-1

Alignment Scores:
Pred. No.: 6,26e-36 Length: 1359
Score: 420.00 Matches: 131
Percent Similarity: 44.20% Conservative: 90
Best Local Similarity: 26.20% Mismatches: 169
Query Match: 17.57% Indels: 110
DB: 1 Gaps: 20

US-09-910-186a-10 (1-450) x US-08-280-228-1 (1-1359)
```

```
Oy 6 AsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsn 25
Db 7 AATCTGATTGTGGTTCATATGAAGAAGATATAGATGTATA----- 51
Oy 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45
Db 52 ---TTAAAAGAGTACAAATTTAAATTTAGATATATAATATTAATATCGATATA 108
Oy 46 SerGlyTyrAsnAlaGluValSerGluGlyAspValGlnLeuAsnProIlePhePro 65
Db 109 TCTGGGTTAAATTCATCTGTATATACATATCCAGATGCTCAATTCGTCGCC----- 159
Oy 66 PheAspPheLysLeuGlySerSerGlyGlu-----AspArgGly 78
Db 160 -----GGAATAAATGCAAGCAATACATTTAGTAACAATGAATCTTCT 204
Oy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98
Db 205 GAAGTTATAGTCATTAACCTATGGATATTTGAATATATATATATGATATGTTAAATTTACC 264
Oy 99 IleSerPheTrpIleArgIleAsnLys---TrpValSerAsnLeuPro----- 113
Db 265 GTTAGCTTTTGGTGGGTTCTCTAAAGTATCTGCTAGTCTATTAGAACAAATATGGCACA 324
```

QY 114 --- GlyTyrThrIleLeuSerValAsnAsn-----SerGlyTyr 127  
 Db 325 AATGAGTATCAATAATAGTCTATGAAACATAGTCTATCAATAGGATCTGGTTGG 384  
 QY 128 SerIleGlyIleSerAsnPheValPheThrLeuLysGlnAsnGluAspSerGlu 147  
 Db 385 AGTGTATCAGTAAAGTATATCTATATGAGCTTAAAGATTCGCGGGAGAGTT 444  
 QY 148 GlnSerIleAsnPheSerTyrAspIleSerAsnAlaProGlyTyr-----AsnLys 165  
 Db 445 AGACAAATACCTTTTAGG--GATTACCTGATATAATCTTATTTAGCAATAAA 501  
 QY 166 TrpPhePheValThrValThrAsnMetMetGlyAsnMetLysIleTyrIleAsnGly 185  
 Db 502 TGGGTTTATTAATCTATCTATGATAGTATCTCTCTAATTTGTATATAAATGCA 561  
 QY 186 LysLeuLeuAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205  
 Db 562 GTACTTATGGGAGTGCAGAAATCTAGTCTTTAGGAGCTATTAGAGGATAATAATA 621  
 QY 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAsnIle 225  
 Db 622 ACATTAACCTAGTAGA-----TGTAATAATAATAATAA 657  
 QY 226 AsnMetIrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn 245  
 Db 658 TACGTTTCTTATGATAAATTTAGGATATTTTGCAGACATTAATCCAAAGAGATTGAA 717  
 QY 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeu 265  
 Db 718 AAATATACACAGTTTATCTATCTATACCTTTTAAAGAGCTCTGGGGAACCTTTA 777  
 QY 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276  
 Db 778 CGATATGATACAGATATTTATTTAATACAGTAGCTTCTAGTTCTAAAGATGTTCAATG 837  
 QY 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289  
 Db 838 AAAAATAACAGATATATATTTGACAAATCGCCATCGCTATCTACTAACGGAATAATG 897  
 QY 290 IleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyrLysIleIle 309  
 Db 898 AATATATATTATAGAGTTATATAAT-----GGACTAAATTTATTATA 942  
 QY 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyGlyAspIleLeuTyr 328  
 Db 943 AAAAGATATACACCTAATAATGAATAGTCTTTTGTAAATCAGGTGATTTTATAAA 1002  
 QY 329 PheAspMetThrIleAsnAsnLysAlaTyr-----AsnLeu 340  
 Db 1003 TTATATGATCATATACATAATATGACACATTTAGTGTATCCGAAAGATGGAATGCC 1062  
 QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
 Db 1063 TTTAATATCTTGTAGAAATCTAAGAGTAGTATTAATGCCCCAGGTATCCCTCTTTAT 1122  
 QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380  
 Db 1123 AAAAATGGAAGCAGTAAA-----TTG 1146  
 QY 381 AsnAsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400  
 Db 1147 CGTGATTTAAACCACTATCTGTACAACTT-----AAATATATCATGATATAAATGCA 1200  
 QY 401 SerGlyIleCysSerIleGlyTyrArgPheArgLeuGlyGlyAsp----- 416  
 Db 1201 TCT---TTAGGACTAGTAGGTACCAATAGTCAATAGGCAACGATCCAAATAGGAT 1257  
 QY 417 -----TrpTyrArgHisAsnTyrLeuValProThrValLysGlnGly 430  
 Db 1258 ATATTAATGCAAGCAACTGTGAC---TTTAATCAATTAAGATAAAT----- 1305

QY 431 AsnTyrAlaSerIleuLeuGluSerThrHisTrpGlyPheValProValSerGlu 450  
 Db 1306 -----TTAGGATGTGATTGGTTACTTTGTACTACAGATGAA 1341  
 RESULT 15  
 US-08-280-228-3  
 ; Sequence 3, Application US/08280228  
 ; Patent No. 5571694  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Makoff Dr, Andrew J  
 ; APPLICANT: Romanos Dr, Michael A  
 ; APPLICANT: Clare Dr, Jeffrey J  
 ; APPLICANT: Fairweather Dr, Neil F  
 ; TITLE OF INVENTION: VACCINES  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: NIXON & VANDERHYE P.C.  
 ; STREET: 1100 No. 5571694th Glebe Road  
 ; CITY: Arlington,  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22201-4714  
 ; COMPUTER READABLE FORM: disk  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/280,228  
 ; FILING DATE: 25-JUL-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/618,312  
 ; FILING DATE: 27-NOV-1990  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 8926832.0  
 ; FILING DATE: 28-NOV-1989  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 9006097.1  
 ; FILING DATE: 17-MAR-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Wilson, Mary J  
 ; REGISTRATION NUMBER: 32,955  
 ; REFERENCE/DOCKET NUMBER: 117-163  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 816-4000  
 ; TELEFAX: (703) 816-4100  
 ; TELEX: 200797 NIXN UR  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1359 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Clostridium tetani  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1356  
 US-08-280-228-3  
 Alignment Scores: 6.26e-36 Length: 1359  
 Pred. No.: 420.00 Matches: 131  
 Score: 44.20% Conservative: 90  
 Percent Similarity: 26.20% Mismatches: 169  
 Best Local Similarity: 17.57% Indels: 110  
 Query Match: 1 Gaps: 20  
 DB:



US-09-910-186A-10 (1-450) x US-08-280-228-3 (1-1359)

```
Oy 6 AsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsn 25
Db 7 AACCTGATTCTTGGTCGCAACGAAGAAGACATGATGTTATC----- 51
Oy 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45
Db 52 ---CTGAAAAAGTCTACCATCTCTGAACCTGGACATCAACAACATATTATTCGCGACATC 108
Oy 46 SerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePro 65
Db 109 TCTGTTTCAACTCCTCTGTTTATCACAATCCAGATGCTCAATGTGCGG----- 159
Oy 66 PheAspPheLysLeuGlySerSerGlyGlu-----AspArgGly 78
Db 160 -----GGCATCAACGGCAAGCTATCCACCTGCTTAACAACGAATCTTCT 204
Oy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98
Db 205 GAAGTTATCGTGCACAAAGGCCATGACATCAACAACAGATGTTCAACAACATTCAC 264
Oy 99 IleSerPheTrpIleArgIleAsnLys---TrpValSerAsnLeuPro----- 113
Db 265 GTTAGCTTCTGGTGGCGTTCGGAAGTTCTGCTTCCACCTGGACAGTACGCGACT 324
Oy 114 ---GlyTyrThrIleIleAspSerValLysAsnAsn-----SerGlyTrp 127
Db 325 AACGAGTACTCCATCATCAGCTCTATGAGAAGAACATCCCTGCTCCATCGGCTCTGGTTGG 384
Oy 128 SerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGlu 147
Db 385 TCTGTTCCCTGAAGGTGTAACAACCTGATCTGGACCTGGAACACTCCGCGCGGCAAGTT 444
Oy 148 GlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr-----AsnLys 165
Db 445 CGTCAGATCACTTTCGCG---GAOCTCGCGACAAAGTTCAACGCTAGCTGGCTAAACAA 501
Oy 166 TrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly 185
Db 502 TGGGTTTTCATCACTATCACTACAGATCGGTCTGTTCTGCTAACCTGTACATCAACGCG 561
Oy 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205
Db 562 GTTCGATGGCTCCGGTGAATCACTGGTCTGGCGCTATCCGTGAGGACAAACAACATC 621
Oy 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIle 225
Db 622 ACTCTTAAGCTGGACCGT-----TGCAACAAACAACAACAC 657
Oy 226 AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn 245
Db 658 TACGTATCCATCGCAAGTTCGGTATCTTCTGCAAGCAGCTGAACCGGAAGATCGAA 717
Oy 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAsnAspLeu 265
Db 719 AAACGTATACAGCTACCTGCTATCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 777
Oy 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276
Db 778 CGTTACGACACCGAATATACCTGATCCGGTAGCTTCTAGCTTAAAGACGTTACGCTG 837
Oy 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289
Db 838 AAAAATCATCACTGATCATGATCACTGACCAACGCGCGCTCTACACTACGCTAAACTG 897
Oy 290 IleValPheAsnThrArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIle 309
Db 898 AACATCTACTACCGAGCTGTACAAC-----GGCCTGAATTCATCATC 942
Oy 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyGlyAspIleLeuTyr 328
Db 943 AAACGCTACTCCGACAAACAAAGAAATCGATCTTCTGTTAAATCTGCTGCTGCTGCTGCTG 1002
```

```
Oy 329 PheAspMetThrIleAsnAsnLysAlaTyr-----AsnLeu 340
Db 1003 CTGTACGTTTCTTTACAAACAACAGAACATCTGTGGTTACCCGAAACAGGTACGCT 1062
Oy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
Db 1063 TTCAACAACCTGGACAGAAATCTCGGTGTGGTTACACAGCTCCGGGTATCCGCTGTAC 1122
Oy 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380
Db 1123 AAAAAAATGGAAGCTGTAA-----CTG 1146
Oy 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Db 1147 CGTGACCTGAAACCTACTCTGTTCAGCTG-----AAACTGTACGACGACAAAAACGCT 1200
Oy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAsp----- 416
Db 1201 TCT---CTGGGTCTGGTGTGTACCCACAGCGGTACGATCGGTAAACGACCCGACCGTAC 1257
Oy 417 -----TrpTyrArgHisAsnTyrLeuValProThrValLysGlnGly 430
Db 1258 ATCCGTGATCGCTTCTTAACCTGTAC---TTCAACACCTGAAGACAAATC----- 1305
Oy 431 AsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheValProValSerGlu 450
Db 1306 -----CTGGGTGCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1341
```

Search completed: November 7, 2002, 19:11:48  
Job time : 77 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 18:38:12 ; Search time 56 Seconds

(without alignments)  
2850.955 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTIFNFTSTNLSLKDI.....NYASILESTHWGFPVSE 450

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Best-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame\_p2n\_model -DEV=xlh  
-cg/cn2\_1/USPTO.spcol/US09910186/runat\_04112002\_111619\_20956/app\_query.fasta\_1.647  
-cg/Published\_Applications\_NA -QFMT=fastap -SUFFIX=rpnb -MINMATCH=0.1  
-CGOPL-0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=DIOSUM62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -SCORE=pcr -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09910186 -CGEN\_1\_1\_18 -runat\_04112002\_111619\_20956  
-ICPU=6 -ICPU=3 -NO\_XLFPY -NO\_MKAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_NA:\*

1:	/cgcn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2:	/cgcn2_6/ptodata/2/pubpna/PTC_NEW_PUB.seq:*
3:	/cgcn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4:	/cgcn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5:	/cgcn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6:	/cgcn2_6/ptodata/2/pubpna/PTCUS_PUBCOMB.seq:*
7:	/cgcn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8:	/cgcn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9:	/cgcn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10:	/cgcn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11:	/cgcn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12:	/cgcn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13:	/cgcn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14:	/cgcn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2381	99.6	1371	10	US-09-350-756-3
2	2364	98.9	3676	12	US-10-051-952-4
3	863.5	36.1	3831	12	US-10-051-952-5
4	588.5	24.6	3894	12	US-10-051-952-9
					Sequence 3, Appl1
					Sequence 4, Appl1
					Sequence 5, Appl1
					Sequence 9, Appl1

5	586.5	24.5	4835	10	US-09-288-326-10	Sequence 10, Appl1
6	585.5	24.5	1400	10	US-09-350-756-4	Sequence 4, Appl1
7	582	24.3	3876	12	US-10-051-952-2	Sequence 2, Appl1
8	581.5	24.3	3753	12	US-10-051-952-6	Sequence 6, Appl1
9	581.5	24.3	3759	12	US-10-051-952-7	Sequence 7, Appl1
10	578.5	24.2	1347	10	US-09-350-756-6	Sequence 6, Appl1
11	568	23.8	2452	10	US-09-350-756-7	Sequence 7, Appl1
12	561	23.5	3876	12	US-10-051-952-3	Sequence 3, Appl1
13	557.5	23.3	1338	10	US-09-350-756-1	Sequence 1, Appl1
14	553.5	23.1	3891	12	US-10-051-952-1	Sequence 1, Appl1
15	550	23.0	1327	10	US-09-350-756-9	Sequence 9, Appl1
16	546	22.8	1293	8	US-08-981-087A-5	Sequence 5, Appl1
17	546	22.8	1313	8	US-08-981-087A-6	Sequence 6, Appl1
18	541	22.6	1341	10	US-09-350-756-2	Sequence 2, Appl1
19	529	22.1	1317	10	US-09-350-756-5	Sequence 5, Appl1
20	516.5	21.6	3825	12	US-10-051-952-8	Sequence 8, Appl1
21	400.5	16.8	1987	10	US-09-350-756-8	Sequence 8, Appl1
22	393	16.4	702	10	US-09-350-756-10	Sequence 10, Appl1
23	168.5	7.0	621	10	US-09-350-756-11	Sequence 11, Appl1
24	145.5	6.1	6025	9	US-09-895-913A-11	Sequence 11, Appl1
25	142.5	6.0	1431	10	US-09-774-414-2	Sequence 2, Appl1
26	133	5.6	2795	10	US-09-815-242-4523	Sequence 4523, Ap
27	133	5.6	2799	10	US-09-815-242-8340	Sequence 8340, Ap
28	129	5.4	2523	10	US-09-815-242-4724	Sequence 4724, Ap
29	129	5.4	2526	10	US-09-815-242-8653	Sequence 8653, Ap
30	128	5.4	4310	10	US-09-740-274-5	Sequence 5, Appl1
31	121.5	5.1	15567	12	US-10-047-676A-3	Sequence 3, Appl1
32	121	5.1	1799	8	US-08-834-666A-19	Sequence 19, Appl1
33	120	5.0	640681	10	US-09-790-988-1	Sequence 1, Appl1
34	118	4.9	2979	10	US-09-815-242-4754	Sequence 4754, Ap
35	118	4.9	3009	10	US-09-815-242-8801	Sequence 8801, Ap
36	118	4.9	2376	10	US-09-815-242-9060	Sequence 9060, Ap
37	116.5	4.9	2376	10	US-09-880-107-3276	Sequence 3276, Ap
38	116.5	4.9	2988	10	US-09-815-242-4196	Sequence 4196, Ap
39	116.5	4.9	3030	10	US-09-815-242-8043	Sequence 8043, Ap
40	116.5	4.9	4389	12	US-10-044-090-611	Sequence 611, App
41	116.5	4.9	4531	10	US-09-925-300-342	Sequence 342, App
42	113.5	4.7	4594	10	US-09-070-927A-230	Sequence 230, App
43	111.5	4.7	1866	10	US-09-856-247A-1	Sequence 1, Appl1
44	111.5	4.7	2367	10	US-09-995-587A-2	Sequence 2, Appl1
45	111.5	4.7	2394	10	US-09-995-587A-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1  
US-09-350-756-3  
; Sequence 3, Application US/09350756  
; Patent No. US20020034521A1  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
; APPLICANT: John S. Lee  
; APPLICANT: Peter Pushko  
; APPLICANT: Michael D. Parker  
; APPLICANT: Jonathan F. Smith  
; APPLICANT: Mark T. Dertzbaugh  
; APPLICANT: Leonard Smith  
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
; FILE REFERENCE: 003/124/SAP RIID 98-21  
; CURRENT APPLICATION NUMBER: US/09/350,756  
; CURRENT FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: US 60/092,416  
; EARLIER FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
; SEQ ID NO 3  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"  
US-09-350-756-3

Alignment Scores:

```

Pred. No.: 7 356-261 Length: 1371
Score: 2381.00 Matches: 448
Percent Similarity: 99.78% Conservativity: 1
Best Local Similarity: 99.56% Mismatches: 1
Query Match: 99.58% Indels: 0
DB: 10 Gaps: 0

US-09-910-186a-10 (1-450) x US-09-350-756-3 (1-1371)

Qy 1 MethThrlProPheAsnLlePheSerTyThrAsnAsnSerLeuLeuLysAspLlelle 20
Db 10 ATCCACCTCCCATCAACATCTCTCTACACCAACAACCTCCCTGTGAAGACATCATC 69
Qy 21 AsnGluTyThrPheAsnAsnLleAsnAspSerLysLleLeuSerLeuGlnAsnArgLysAsn 40
Db 70 AACGAGTACTTCACACATCAACAGCTCCAGATCTCTGCTCCGCGAGAACCGTAAAGAC 129
Qy 41 ThrLeuValAspThrSerGlyTyThrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60
Db 130 ACCCTGGTCCGACACCTCCGCTTACACGCGGAGGTCTCCGAGGAGGTGACGTCCAGCTG 189
Qy 61 AsnProLlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80
Db 190 AACCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCGCGTGGAGACAGAGTAAAGTC 249
Qy 81 IleValThrGlnAsnGluAsnLleValTyThrAsnSerMetTyThrGluSerPheSerLleSer 100
Db 250 ATCGTCAACCAAGAGAGACATCGTCTACAACTCCATGTCATGTCAGAGTCTTCCATCTCC 309
Qy 101 PheTrpLleArgLleAsnLysTyThrValSerAsnLeuProGlyTyThrLleLleAspSer 120
Db 310 TTCTGGATCAACATCAACAGTGGTCTCCAACTTCCAGCTTACACCATCATCGACTCC 369
Qy 121 ValLysAsnAsnSerGlyTyThrSerLleGlyLleLleSerAsnPheLeuValPheThrLeu 140
Db 370 GTCAAGAACTCCGCTGGTCTCCATCGTATCATCTCCACCTTCCTGGTCTTCAACCTG 429
Qy 141 LysGlnAsnGluAspSerGluLysSerLleAsnPheSerTyThrAspLysSerAsnAla 160
Db 430 AAGCAAGACGAGGACTCCGAGCTCCATCACTTCTCCAGCATCTCCCAACACGCT 489
Qy 161 ProGlyTyThrAsnLysTyThrPheValThrValThrAsnAsnMetMetGlyAsnMetLys 180
Db 490 COTGTTTACCAAGTGTCTTCTCGTCAACCTCACCACCAACATGATGGTAAACATGAAG 549
Qy 181 IleTyThrLleAsnGlyLysLeuLleAspThrLleLysValLysGluLeuThrGlyLleAsn 200
Db 550 ATCTACATCAACAGTAAAGCTGATCGACACCATCAAGTCAAGGAGTGAACGCTATCAAC 609
Qy 201 PheSerLysThrLleThrPheGluLleAsnLysLleProAspThrGlyLleLleThrSer 220
Db 610 TTCTCCAGACCATCACTTCGAGATCAACAGATCCACACACCGGTCTGATCACTCC 669
Qy 221 AspSerAspAsnLleAsnMetTrpLleArgAspPheTyThrLlePheAlaLysGluLeuAsp 240
Db 670 GACTCCGACATATCAACATGCTGCTGCTTCACTTCACTTCCGCGAGGAGTGGAC 729
Qy 241 GlyLysAspLleAsnLleLeuPheAsnSerLeuGlnTyThrAsnValLysAspTyThr 260
Db 730 GGTAAAGACATCAACATCTCTTCACTTCTGAGTACACCAAGCTCGTCAAGGACTAC 789
Qy 261 TrpGlyAsnAspLeuArgTyThrAsnLysGluTyThrMetValAsnLleAspTyThrLeuAsn 280
Db 790 TGGGTACGACGCTGAGATACAAACAGGAGTACTTACATGCTCAACATCGACTACTTGAAC 849
Qy 281 ArgTyMetTyThrAlaAsnSerArgGlnLleValPheAsnThrArgArgAsnAsnAsp 300
Db 850 AGATACATGTACGCGCACTCCAGACAGATGCTCTTCAACACCAAGACGCTAACACACAGAC 909
Qy 301 PheAsnGluGlyTyThrLysLleLleLleLysArgLleArgGlyAsnThrAsnAspThrArg 320
Db 910 TTCAACGAGGGTTTAAAGATCATCATCAAGGTTATCAGAGGTAAACACCAACGACACGAG 969

```

Copied from 10910186 on 05-05-2004

```

Qy 321 ValArgGlyGlyAspLleLeuTyThrPheAspMetThrLleAsnAsnLysAlaTyThrAsnLeu 340
Db 970 GTGAGAGGTGGTACATCTCTGACTTCGACATGACTATCAACAAACAGGCTTACACCTG 1029
Qy 341 PheMetLysAsnGluThrMetTyThrAlaAspAsnHisSerThrGluAspLleTyThrAla 360
Db 1030 TTCAATGAAGAACAGAGACCATGTACGCGCAACACCATCCACGAGGACATCTACGCCATC 1089
Qy 361 GlyLeuArgGluGlnThrLysAspLleAsnAspAsnLleLlePheGlnLleGlnProMet 380
Db 1090 GGTCTGGTGGAGCAGACACAGGACATCAACAGCAATCTTCCAGATCCAGGCAATG 1149
Qy 381 AsnAsnThrTyThrTyThrAlaSerGlnLlePheLysSerAsnPheAsnGlyGluAsnLle 400
Db 1150 AACAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1209
Qy 401 SerGlyLysSerLleGlyTyThrArgPheAspLeuGlyLysTyThrArgHis 420
Db 1210 TCCGGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1269
Qy 421 AsnTyThrLeuValProThrValLysGlnGlyAsnTyThrAlaSerLeuGluSerThrSer 440
Db 1270 AACTACTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1329
Qy 441 ThrHisTrpGlyPheValProValSerGlu 450
Db 1330 ACCCATGGGATTCGTCCTCCAGTCTCCGAG 1359

RESULT 2
US-10-051-952-4
; Sequence 4, Application US/10051952
; Patent No. US20020107199A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin
; FILE REFERENCE: 2933CIP
; CURRENT APPLICATION NUMBER: US/10/051,952
; PRIOR FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 09/730,237
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3876
; TYPE: DNA
; ORGANISM: botulinum toxin
US-10-051-952-4

Alignment Scores:
Pred. No.: 2 8e-258 Length: 3876
Score: 2364.00 Matches: 447
Percent Similarity: 99.55% Conservativity: 0
Best Local Similarity: 99.55% Mismatches: 2
Query Match: 98.87% Indels: 0
DB: 12 Gaps: 0

US-09-910-186a-10 (1-450) x US-10-051-952-4 (1-3876)

Qy 2 ThrLeProPheAsnLlePheSerTyThrAsnAsnSerLeuLeuLysAspLleLleAsn 21
Db 2527 ACAATACCTTTAATATTTTTCATATATATATTTTTCATATATATATTTTTCATATATAT 2586
Qy 22 GluTyThrPheAsnLleAsnAspSerLysLleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2587 GAATATTTCAATATATATATATATATATTTTTCATATATATATTTTTCATATATATAT 2646
Qy 42 LeuValAspThrSerGlyTyThrAsnAlaGluValSerGluGlyAspValGlnLeuAsn 61
Db 2647 TTAGTGATATATATATATATATATATATATATATATATATATATATATATATATAT 2706
Qy 62 ProLlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValLle 81

```

Db 2707 CCAATATTTCATTTGACTTTAAATTTAGTAGTTTCAGGGAGGATAGAGTAAAGTTATA 2766  
Qy 82 ValThrGlnAsnGluAsnIleValTyArgSerMetTyArgSerPheSerIleSerPhe 101  
Db 2767 GTACCCAGAAATGAATATATGATATATTAATCTGATGATGAAAGTTTATGCAATAGTTT 2826  
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyThrIleIleAspSerVal 121  
Db 2827 TCGATTAGATAAATAAATGGTAAAGTAATTTACCTCGATATACTATAATTGATAGTT 2886  
Qy 122 LysAsnAsnSerGlyTyArgSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
Db 2887 AAAATAAATCACTGAGTATAGGATATATATGATATATTTTATGATATTTTAAATA 2946  
Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyArgSerAsnAsnAlaPro 161  
Db 2947 CAAAATGAAGATAGTGAACAAAGTATAAATTTAGTTATCATATATCAAAATATGCTCT 3006  
Qy 162 GlyTyArgAsnLysTrpPheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
Db 3007 GGATAGAATAAATGGTTTGTAACTGTACTAACAAATATGATGGAATATGAAGATT 3066  
Qy 182 TyRileAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
Db 3067 TATATAATGGAATTAATAGATCTATAAAGTTAAAGAACTAACTGGAATTAATTT 3126  
Qy 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221  
Db 3127 ACCAAATATATAAATTTGAATTAATAAATCCAGATACCGTGTGATTAATCTCAGAT 3186  
Qy 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyRilePheAlaLysGluLeuAspGly 241  
Db 3187 TCTGATACATCAATATGCTGGATAGAGATTTTATATATTTGCTAAAGAAATAGATGT 3246  
Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValValLysAspTyTrp 261  
Db 3247 AAGATATATATATATTTAAATAGTTCGAATATATACTAATGTTGTAAGAAGATTATGG 3306  
Qy 262 GlyAsnAspLeuArgTyArgLysGlyTyArgMetValAsnIleAspTyRileAsnArg 281  
Db 3307 GGAATGATTTAAGATATATAAAGATATTTATGTTTATATAGATTTATTAATATAGA 3366  
Qy 282 TyrMetTyRalaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPhe 301  
Db 3367 TATATGATGCAACTACGACAAATTTGTTTAAATACACGATAGAAATATATGACTTC 3426  
Qy 302 AsnGluGlyTyRileIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
Db 3427 AATGAGGATATAAATAATTAATAAAGAAATCAGAGAAATACAAATGATAGTAGTA 3486  
Qy 322 ArgGlyAspIleLeuTyRArgPheAspMetThrIleAsnAsnLysAlaTyArgLeuPhe 341  
Db 3487 CGAGGAGGAGATATTTATATTTGATGATGACAATTAATAACAAGCATATAATTTGTT 3546  
Qy 342 MetLysAsnGluThrMetTyRalaAsnAsnHisSerThrGluAspIleTyRAlaIleGly 361  
Db 3547 ATGAGAATGAACATATGATGCAATATATCATAGTACTGAGATATATATGCTATAGGT 3606  
Qy 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
Db 3607 TTAAGNAGACAACAAGGATATAAATGATATATATTTTCAATACACCAATGAAT 3666  
Qy 382 AsnThrTy 401  
Db 3667 AATACCTTATTTAGGCATCTCAAAATATTTAAATCAAAATTTTAAATGAGAAATATTTCT 3726  
Qy 402 GlyIleCysSerIleGlyTyArgPheArgLeuGlyGlyAspTrpTyArgHisAsn 421  
Db 3727 GGAATGTTTCAATGAGTACTTATCTGTTTACCTGAGCTGAGTGGTATAGACCAAT 3786  
Qy 422 TyrLeuValProThrValLysGlnGlyAsnTyRAlaSerLeuLeuGluSerThrSerThr 441  
Db 3787 TATTTGCTGCTACTGTGAGCAAGAAATTTGTTTATTTATTTATTTATTTATTTATTTATTT 3846

Qy 442 HisTrpGlyPheValProValSerGlu 450  
Db 3847 CATGGGGTTTGTACCTGTAGTGAA 3873  
RESULT 3  
US-10-051-952-5  
; Sequence 5, Application US/10051952  
; Patent No. US20020107199A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Patricia  
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
; FILE REFERENCE: 2933CIP  
; CURRENT APPLICATION NUMBER: US/10/051,952  
; CURRENT FILING DATE: 2002-01-17  
; PRIOR APPLICATION NUMBER: 09/730,237  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 3831  
; TYPE: DNA  
; ORGANISM: botulinum toxin  
US-10-051-952-5  
Alignment Scores:  
Pred. No.: 4,71e-88 Length: 3831  
Score: 863.50 Matches: 136  
Percent Similarity: 61.20% Conservative: 80  
Best Local Similarity: 43.46% Mismatches: 148  
Query Match: 36.11% Indels: 27  
Gaps: 11  
US-09-910-186a-10 (1-450) x US-10-051-952-5 (1-3831)

Qy 2 ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
Db 2515 ACAATGCGCTTTTAAATTTTTCATATCTAATAATAATTTCTTTATTAAGATATAATAAT 2574  
Qy 22 GluTyRArgAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
Db 2575 GAATATTTCAATAGTATTAATGATTCAAAATTTTGAGCTTACAAAACAAAAAATGCT 2634  
Qy 42 LeuValAspThrSerGlyTyArgAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
Db 2635 TTAGTGGATACATCAGGATATAATGACAGAGTGGAGGTAGGAGATATAATGCTCAACTAAT 2694  
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
Db 2695 AGGATATATACAAATGACTTTAAATTAAGTAGTTACAGAGAT-----AAATATATA 2745  
Qy 82 ValThrGlnAsnGluAsnIleValTyArgSerMetTyRArgSerPheSerIleSerPhe 101  
Db 2746 GTAAATTTTAAATAATAATATTTTATATAGCGCTATTTATGAGAACTCTAGTGTAGTTT 2805  
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyrThrIleIleAspSer 120  
Db 2806 TCGATTAGATATCTTAAAGATTTTAACTAAATTTCTCATATGAATGAATATAAATAAATACGT 2865  
Qy 121 ValLysAsnAsnSerGlyTyRArgSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140  
Db 2866 ATAGACAAATAATCTGGGTGGAATTTATGATAGGAAATGCAATATAGATGATTTTA 2925  
Qy 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyArgSerIleSerAsnAla 160  
Db 2926 CAAGATGTTTAAATAGAAAGTATAAAGTTTAAATTTTATGATGATGATTAAGTAT 2985  
Qy 161 ProGlyTyR---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet 179  
Db 2986 ACAGATATACAAATTAATGTTTCTTACTATATACTAATAATATAATGAGGATATG 3045  
Qy 180 LysIleTyRileAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199  
Db 180 LysIleTyRileAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199

Db 3046 AACCTTTATATAATGGAGAAATAAGACAGAGTCAAAATAATTGAAGATTAGATGAGGTT 3105  
Qy 200 AsnPheserLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr 219  
Db 3106 AAGTTAGATAAACCATTAGTATTGGATAGAGTGAAGAAATATAGAT----- 3150  
Qy 220 SerAspSerAsnIleAsnMetTrpIleArgAspPheTyIlePheAlaLysGluLeu 239  
Db 3151 -----GAGAATCAGATGCTTTGGATTAGAGATTTTAATATTTTCTTAAGAATA 3201  
Qy 240 AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValValLysAsp 259  
Db 3202 AGTAATGAAGATATTAATATTGATATGAGGGGACAAATATTGAAGAAATGTTATTAAAGAT 3261  
Qy 260 TyrTrpGlyAsnAspLeuArgTyAsnLysGluTyTyMetValAsnIleAspTyIle 279  
Db 3262 TATTGGGAAATCCTTTGAAGTTGATACAGATAATATATATTAATGATTAATATATA 3321  
Qy 280 AsnArgTyMetTyAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn 299  
Db 3322 CATAGGTATATGCACCTGAAGCTAATGTACTTGTACTTGTTCGGTATCCAGATAGATCT 3381  
Qy 300 AspPheAsnGluGlyTyIleLysIleLeuLysArgIleArgGlyAsnThrAsnAspThr 319  
Db 3382 AAATATATATCTGGAATCCTATTAATCTTAATCAATGATCTGATCAAGATCCTTATAGT 3441  
Qy 320 ArgValArgGlyAspIleLeuTyPheAspMetThrIleAsnAsnLysAlaTyAsn 339  
Db 3442 AGAATTTTAAATGAGAGATAATATAATCTTCATATCTTATATAATAGTAGGAATATATG 3501  
Qy 340 LeuPheMetLysAsnGluThrMetTyAlaAspAsnHis-----SerThrGluAsp 356  
Db 3502 ATAATAGAGATATCTGATCAATATATGACACACAGGAGGAGATGTTCAACAAATGTT 3561  
Qy 357 IleTyAlaIleGlyLeuArgGluGlnThrLysAspLysAsnAsnIlePheGln 376  
Db 3562 GTATATGCAATTAATACAGAGTAATTTAGTAATATATGTTAGTATAGT---ATATTAGT 3618  
Qy 377 IleGlnProMetAsnAsnThrTyTyTyAlaSerGlnIlePheLysSerAsnPheAsn 396  
Db 3619 ATAAAAATATGATCTCAAAAAATAATATTTGATCAAAATTTTC---TCTAGTTTAGG 3675  
Qy 397 GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyArgPheArgLeuGlyGly 415  
Db 3676 ---GAAATACATAGCTCTAGCAGATATATATAAACTTGGAGATTTTCTTTT----- 3726  
Qy 416 AspTrpTyArgHisAsnTyLeuValProThrValLysGlnGlyAsnTyAlaSerLeu 435  
Db 3727 -----AAAAATGCATAGACGCCA---GTTGCAGTAACTAATATATGAACAAAA 3771  
Qy 436 LeuGluSerThrSerThrHisTrpGlyPheVal 446  
Db 3772 CTATTATCACTTCATCTTTTGGAAATTTATT 3804

## RESULT 4

US-10-051-952-9  
; Sequence 9, Application US/10051952  
; Patent No. US20020107199A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Patricia  
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
; FILE REFERENCE: 2933CIP  
; CURRENT APPLICATION NUMBER: US/10/051,952  
; CURRENT FILING DATE: 2002-01-17  
; PRIOR APPLICATION NUMBER: 09/730,237  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 3894  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: primers used to introduce Stu I and EcoR I  
; OTHER INFORMATION: restriction sites into the 5' and 3' ends of the  
; OTHER INFORMATION: BONY/A-L chain gene fragment  
US-10-051-952-9

Alignment Scores: 7.63e-57 Length: 3894  
Pred. No.: 588.50 Matches: 155  
Score: 49.39% Conservative: 87  
Percent Similarity: 31.63% Mismatches: 171  
Best Local Similarity: 24.61% Indels: 77  
Query Match: 12 Gaps: 17  
Db:

US-09-910-186a-10 (1-450) x US-10-051-952-9 (1-3894)

Qy 2 ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
Db 2518 AGTATACCATTTGATCTTTCACATATATACCAAGGACACAAATTTTAAACAAGTTTAAAT 2577  
Qy 22 GluTyPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGluAsnArgLysAsnThr 41  
Db 2578 AATTATATTAGTATATATAGTAGTATCTATTTAAAGTTTAAGTTATAGAGGTGGCGT 2637  
Qy 42 LeuValAspThrSerGlyTyAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
Db 2638 TTTATAGATTCACTCTGGATATGTCGCACTATGATGATGATGATGATGATGATGATGATGAT 2697  
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyLysArgGlyLysValIle 81  
Db 2698 GATATAGGAATGCTCAATTTAAATTAATAATTCG-----GAAATAGTAAATATAGC 2751  
Qy 82 ValThrGlnAsnGluAsnIleValTyAsnSerMetTyTrpSerPheSerIleSerPhe 101  
Db 2752 GCACATCAAGTAAATTCGTTGATATGATGATGATGATGATGATGATGATGATGATGAT 2811  
Qy 102 TrpIleArgIleAsnLysTrpValSerAsn-----LeuProGlyTy 115  
Db 2812 TGGTAAAGGACTCTCAATATAATAATAATATGATATACAACTTATCTTCAAAATGAGTAT 2871  
Qy 116 ThrIleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPhe 135  
Db 2872 ACAATAATTAGTTGATATAAAATGACTCAGGATGGAAGTATCTATTAAAGGAATAGA 2931  
Qy 136 LeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyAsp 155  
Db 2932 ATAATATGGACATTAATAGATGTTAATGCAAAATCTAAATATATTTTCGAATATAGT 2991  
Qy 156 IleSerAsnAsnAlaProGlyTy---AsnLysTrpPhePheValThrValThrAsnAsn 174  
Db 2992 ATAAAGATATATATCAGATTAATATAATAAAGTTTCCATACTATTACTAATGAT 3051  
Qy 175 MetMetGlyAsnMetLysIleTyIleAsnGlyLysLeuIleAspThrIleLysValLys 194  
Db 3052 AGATTAGTAAACGCAATATTTATATAAATGGAAGTTTGAAAAAAGTCAAAAAATTTTA 3111  
Qy 195 GluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAsp 214  
Db 3112 AACTTAGATAGATAATTTCTAGTAATGATATAGACTCAATTAATTAATTAATTAATTAAT 3171  
Qy 215 ThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyIle 234  
Db 3172 ACTACTAAATTT-----GTTTGGATTAAGGATTTTAAATATT 3207  
Qy 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyThr 254  
Db 3208 TTTGGTAGAGAAATTAATGCTACAGAAATATCTTCACTATATATTTGGATTCATCATCTACA 3267  
Qy 255 AsnValValLysAspTyTrpGlyAsnAspLeuArgTyAsnLysGluTyTyMetVal 274  
Db 3268 AATACTTTAAAGATTTTGGGGAATCCCTTTAAGATAGGATACACAAATACTATCTGTT 3327  
Qy 275 AsnIleAspTyLeuAsnArgTyTyMetTyAlaAsnSerArgGlnIleValPheAsnThr 294

```

3328  RATCAAGCTATGCCAAATATCTATATAAAGTATTTTACTAAAGCTTCTATGCGGGAAC 3387
QY 295  ArgArgAsnAsnAsnAspPheAsnGlu-----Gly 304
Db 3388  GCACACGACACAAAC--TTTAATAATGCAGCAATAAATATCAAAATTTATATCTTGGT 3444
QY 305  TyrLysIleIleIleLysArgIle-----ArgGlyAsnThrAsnAspThrArgVal 321
Db 3445  TTACAGTATTATATAAAAAAGCATCAAAATCTCGAATATAAATAATAGATAATATAGTC 3504
QY 322  ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341
Db 3505  ACAGAAGGAGAGATATATATATCTTAATATT----- 3534
QY 342  MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAsp-----IleTyrAla 359
Db 3535  -----CATAAATATTTCTGATGAATCTTAGAGAGATATATGTT 3570
QY 360  IleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPro 379
Db 3571  TTG-----GTCAATCTTAAGCAATTCAAACTCAATATT-----TTAGCACCC 3615
QY 380  MetAsnAsn-----ThrTyrTyrTyrAlaSerGlnIle-----PheLysSer 393
Db 3616  ATAAATGATGATCCCTACCTTCTATGATGATGATGATGATGATGATGATGATGATGATG 3675
QY 394  AsnPheAsnGlyGluAsnIleSer-----GlyIleCysSerIle 406
Db 3676  ACATATAATGTCAGATACCTTTGGCAAAAGATACATAAACAATTTGGCGCTTTGGNATT 3735
QY 407  GlyThrTyrArgPheArgLeuGlyGlyAspTyrPyrArgHis-----AsnTyrLeuValPro 425
Db 3736  GCTAAATTTGTTAAAGATTATGATATCTTTGGGATACCTATGATGATGATGATGATGATG 3789
QY 426  ThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 3790  TGCATATGCTAGCTGATCTCGAAGATATCTCGAATAATATAAATAAATTAAGCTGGGA 3849
QY 441  ThrHisTrpGlyPheValProValSerGlu 450
Db 3850  TGTAAATGGCAATTCATCCCGTGGATGAA 3879

```

## RESULT 5

us-09-288-326-10

Sequence 10, Application US/09288326

Patent No. US20010018049A1

GENERAL INFORMATION:

APPLICANT: Kei Roger Aoki

APPLICANT: George Sachs

TITLE OF INVENTION: Method and Compositions for the

FILE OF INVENTION: Treatment of Pancreatitis

FILE REFERENCE: 17282

CURRENT APPLICATION NUMBER: US/09/288,326

CURRENT FILING DATE: 1999-04-08

NUMBER OF SEQ ID NOS: 11

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 10

LENGTH: 4835

TYPE: DNA

ORGANISM: Clostridium botulinum

us-09-288-326-10

## Alignment Scores:

Pred. No.:	1,76e-56	Length:	4835
Score:	586.50	Matches:	149
Percent Similarity:	48.94%	Conservative:	82
Best Local Similarity:	31.57%	Mismatches:	186
Query Match:	24.53%	Indels:	55
DB:	10	Gaps:	11

us-09-910-186a-10 (1-450) x us-09-288-326-10 (1-4835)

```

QY 3  IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22
Db 2902  ATACCTTTTCAGCTTTCCAAATACGTAGATAAATCAAGATATATATCTACATTTACTGAA 2961
QY 23  TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
Db 2962  TATATTAAAGTAATATTATTAATCTCTATATTAATTTAAAGATATGAAGTAATATTTA 3021
QY 43  ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62
Db 3022  ATAGACTTATCTAGTATGTCATCAAAATTAATATTGGTAGTAAGTAATTTTGATCCA 3081
QY 63  IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79
Db 3082  ATAGATAAAATCAAAATCAATTTATTTAATTTAGAAAGTAGT-----AAA 3126
QY 80  ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99
Db 3127  ATTGAGGTAATTTTAAATAATGCTATTGTTATATAAATAGTATGATGAAAATTTTAGTACT 3186
QY 100  SerPheTyrPheArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116
Db 3187  AGCTTTTGGATGAAGATTTCTTAAGTATTTTAACAGATTAAGTCTAAATAATGAATATACA 3246
QY 117  IleLeuAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleLeuSerAsnPheLeu 136
Db 3247  ATAATAAATTCATGGAATAATTAATTCAGATGGAAGTATCACTTAATTAATGTTGTAATA 3306
QY 137  ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
Db 3307  ATCTGAGCTTTTACAGGATCTCAGGAATAAACAAGAGTAGTTTTTAAATACATCTCAA 3366
QY 157  SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
Db 3367  ATGATTATATATATCATGCTAGTAGTAATAATATATGTTTAAATAGATGTTGTAGAGATA 3426
QY 176  MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuLeuAspThrIleLysValLysGlu 195
Db 3427  TTAAATAACTCTAAATTTATATAATGAGAGATTAATATATCAATCAAAACCAATTTCAAAT 3486
QY 196  LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
Db 3487  TTAGGTAATATTCATGCTAGTAGTAATAATATATGTTTAAATAGATGTTGTAGAGATA 3546
QY 216  GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235
Db 3547  -----CATAGATATATTGGATAAAATATTATTAATCTTTTT 3582
QY 236  AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255
Db 3583  GATAAGGAATTAATGAGAAAAGAAATCAAGATTTATATGATAATCAATCAAAATTCAGGT 3642
QY 256  ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275
Db 3643  ATTTTAAAGAGACTTTGGGTGATTTTACAATATGATAAACCATCTATATGTTAAAT 3702
QY 276  IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287
Db 3703  TTATATGATCCAAATAATATGTCGATGTAATAATAATAGGTATTAGAGGTATATGTA 3762
QY 288  -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302
Db 3763  CTAAAGGGCCTAGAGGTAGCGTAAATGACTACAAACATTTTAAATCAAGTTTCTAT 3822
QY 303  GluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322
Db 3823  AGGGGACAAAATTTATTTATAAATAATATCTCTCGAATAAAGATAATATTGTTAGA 3882
QY 323  GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342
Db 3883  AATAATGATCGGTATATATTAATAGTAGTAGTTAAATAAAGAAATATAGCTTA----- 3936
QY 343  LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362

```

Db 3937 -----GCTACTAATGCACAGCGGCTAGAAAAATACTAAGT 3978  
 QY 363 ArgGluGlnThrLysAspIleAsnAspAsnIlePheGlnIleClnPProMetAsnAsn 382  
 Db 3979 GCATTAGAAATACCTGATGAGAAATCTAAGTCAAGTAGTAGTAAAGTCAAAAAAT 4038  
 QY 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402  
 Db 4039 GATCAAGGAATAACAATAATGC--AAATGAATTTACAAGATAATAATGGGAATGAT 4095  
 QY 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyr 422  
 Db 4096 ATAGCGTTTATAGATTTTCATGAGTT-----AATAAT 4128  
 QY 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440  
 Db 4129 ATACCTAACTAGTAGCAAGTAATTTGGTATAATAGACAAATAGAAAGATCTAGTAGGACT 4188  
 441 -----ThrHisTrpGlyPheValProValSerGlu 450  
 4189 TTGGGTGCTCATGGGAATTTATCTCTGTAGATGAT 4224

## RESULT 6

US-09-350-756-4

Sequence 4, Application US/09350756

Patent No. US20020034521A1

## GENERAL INFORMATION:

APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases

APPLICANT: John S. Lee

APPLICANT: Peter Pushko

APPLICANT: Michael D. Parker

APPLICANT: Jonathan F. Smith

APPLICANT: Mark T. Dertzbaugh

APPLICANT: Leonard Smith

TITLE OF INVENTION: Botulinum Neurotoxin Vaccine

FILE REFERENCE: 003/124/SAP RIID 98-21

CURRENT APPLICATION NUMBER: US/09/350,756

CURRENT FILING DATE: 1999-07-09

EARLIER APPLICATION NUMBER: US 60/092,416

EARLIER FILING DATE: 1998-07-10

NUMBER OF SEQ ID NOS: 11

SOFTWARE: Apple Macintosh Microsoft Word 6.0

SEQ ID NO 4

LENGTH: 1400

TYPE: DNA

ORGANISM: Artificial

FEATURE:

OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"

US-09-350-756-4

## Alignment Scores:

Seq. Id. No.	Length:	3.81e-57	1400
Score:	585.50	Matches:	152
Percent Similarity:	51.72%	Conservative:	89
Best Local Similarity:	22.62%	Mismatches:	160
Query Match:	65	Indels:	65
DB:	10	Gaps:	17

US-09-910-186a-10 (1-450) x US-09-350-756-4 (1-1400)

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnSerLeuLeuLysAspIleIleAsn 21  
 Db 64 AGTATCTCTTTAAGCTTCTCTTATACAGATGATAAAATTTAATTCCTACTTCAAC 123  
 QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
 Db 124 AAGTCTCTCAAGAGAAATTAAGTCTCTTCCTGTTTAAACATGAGATACAAAGATGATAAA 183  
 QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
 Db 184 TAGCTCGACACTTCGGGTAGACTCCCAATATCAACATTAACGCTGACGTGACAAAGTAC 243

QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
 Db 244 CCAACTAACAACAAACCAATTC-----GCTATCTACAACGACAGCTTACTGAGGTGAAC 297  
 QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
 Db 298 ATCTCTCAAAACGACATACATTATCTACGACAAACAAGTACAAGAACTTCCTATTTCCTTC 357  
 QY 102 TrpIleArgIle-----AsnLysTrpValSerAsnLeuProGlyTyrThrIle 117  
 Db 358 TGGGTCAAGGATTCCTAACTACGACAAACAAGATCGTCAACGGTTAAACAAGGAGTACACTATC 417  
 QY 118 IleAspSerValLys---AsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeu 136  
 Db 418 ATCAACGTATGAGACACAACTCCGTTTGAAGTCTCTCTTAACCAACGACGAGATC 477  
 QY 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 478 ATTGGGACCTTGAACGACAAACGAGGTATTAAACAAAAGTTAGCATTCAACTACGCTTAAC 537  
 QY 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
 Db 538 GCAACGGTATTTCTGACTACATCAACAAGTGGATTTTCGTCRATATCACTAACGACAGA 597  
 QY 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195  
 Db 598 TTAGTGACTCTAAGCTTTTACATTAAACGGTAACTTAATCGACAAAAGTCCATTTTAAAC 657  
 QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215  
 Db 658 TTAGTAAACATTTCAGCTTCTGACAACTTTATTCAAGATCGTTAACTGCGATTACACC 717  
 QY 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235  
 Db 718 AGATACATT-----GGCATTAGATACCTCAACATTTTC 750  
 QY 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255  
 Db 751 GACAGGAGTTAGACGAGACCGAGATTCAACTTTATACGACAAAGCACTAACCCAAAT 810  
 QY 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275  
 Db 811 ATTTTGAGGACTTCGGGGTAACTACTGCTTTACGACAAAGGAATACTACTATTAAAC 870  
 QY 276 IleAspTyrLeuAsnArgTyrMet-----TyrAlaAsnSerArgGlnIleValPheAsn 293  
 Db 871 GTGTTAAAGCCAAACAACTTCATTGATGAGAGAAAGGATTCCTTTAAGCATTAACAAC 930  
 QY 294 ThrArgArgAsn-----AsnAsnAspPheAsnGluGlyTyrLysIleIleLys 310  
 Db 931 ATCAGAGCAGCTATCTTTAGCTACAGATTATATCTCTGGTATCAAGGTTAAGATCCAA 990  
 QY 311 ArgIle---ArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPhe 329  
 Db 991 AGAGTTAAACAACTCTCTTACTAACGATACTTGTAGAAAGAACGATCAGGTCTATAT 1050  
 QY 330 AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGlnThrMetTyrAla 349  
 Db 1051 AACTTCGCTGCTAGCAAGACT-----CACTTATTC-----CCATTATATGCT 1092  
 QY 350 AspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIle 369  
 Db 1093 GATACCGCTACCAACCAAC-----AAGGAGAAGACCATCAAGATC 1131  
 QY 370 AsnAsp-----AsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAla 387  
 Db 1132 TCCCTCCTCGGCAACAGATTAAACCAATCGGTGCTGATGAATCCCTCGGTACAACTGT 1191  
 QY 388 SerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGly 407  
 Db 1192 ACCATGACCTTAAATAAT---AATGGAATAAT-----ATTGGG 1230  
 QY 408 ThrTyrArgPheArg-----LeuGlyGlyAspTyrTyrArgHisAsnTyrLeu 423

Db 1231 TGTGTAGTTTCAGGCGAGATAGTGTAGTTCGTAGTCTGGTAT 1275  
 Qy 424 ValProThrValysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrHis 442  
 Db 1276 -----TATACCCACATGAGATACACACACAGCAATGGA 1311  
 Qy 443 -----TrpGlyPheVal 446  
 Db 1312 TGTGTTTGGAACTTTATT 1329

## RESULT 7

US-10-051-952-2  
 ; Sequence 2, Application US/10051952  
 ; Patent No. US20020107199A1  
 ; GENERAL INFORMATION:

APPLICANT: Walker, Patricia

TITLE OF INVENTION: Methods of Administering Botulinum Toxin

CURRENT FILING DATE: 2002-01-17

PRIOR FILING DATE: 2000-12-05

NUMBER OF SEQ ID NOS: 12

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 3876

TYPE: DNA

ORGANISM: botulinum toxin

US-09-910-186a-10 (1-450) x US-10-051-952-2 (1-3876)

Alignment Scores:

Seq. No.: 4,14e-56 Length: 3876  
 Score: 582.00 Matches: 161  
 Percent Similarity: 51.33% Conservative: 89  
 Percent Local Similarity: 33.06% Mismatches: 163  
 Query Match: 24.34% Indels: 74  
 Gaps: 12

3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsnGlu 22  
 2506 ATACCTTTGATCTTCACTGATCTCACTATCTCAATATGCAATATCTATAAAATATTATAAA 2565  
 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42  
 2566 TATAATACGGAATTTTAAATATATATCTTAAATTTAAGATATAGAGATAATAATTTA 2625  
 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluAspValGlnLeuAsnPro 62  
 2626 ATAGATTTATCAGGATGAGCAAGGTAGAGTATATATGCGGTCAAGCTTAATGAT 2685  
 63 IlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleVal 82  
 2686 -----AAAAATCAATTTAAATTAACCTAGTTCACAGAT-----AGTAAGATTAGAGTC 2733  
 83 ThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrp 102  
 2734 ACTCAAAATCAGATATATATTTAATAGATGTCCTCTGATTTTACGGTATAGCTTTTG 2793  
 103 IleArgIleAsnLysTrp-----ValSerAsnLeu-----ProGlyTyrThr 116  
 2794 ATAAGATACCTAAATATAGGAATGATGATATACAAATATATATTCATAATGATATACG 2853  
 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
 2854 ATAATTAATTTATGATAAAATTAATTCAGGTGGAATATCTATTAGGGGTAATAGGATA 2913  
 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 2914 ATATGACCTTAATTTGATATATAATGGAACCAACCAATCAGTATTTTGTGATATACATA 2973

## RESULT 8

US-10-051-952-6

; Sequence 6, Application US/10051952

; Patent No. US20020107199A1

; GENERAL INFORMATION:

Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
 Db 2974 AGAAGATATATACAGAGTATATAAAATAGATGTTTTTGTAACTATTACTATAAT--- 3030  
 Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValysGlu 195  
 Db 3031 TTGGATATGCTAAATTTATATTAATGGCAGTTAGAAATCAATATGATATTAAGAT 3090  
 Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215  
 Db 3091 ATAGGAGAAGTTTATGTTAATGGTGAATAAATCAATTTAAATTAGAT----- 3135  
 Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsn---MetTrpIleArgAspPheTyrIle 234  
 Db 3136 -----GGTGTGTAGTAGACACACAATTTATTTGGATGAATATTTTAGTATT 3183  
 Qy 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThr 254  
 Db 3184 TTTAATACGCAATTAATCAATCAATATATAAAGATATATAAAATTTCAATCATATAGC 3243  
 Qy 255 AsnValLysLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTrpMetVal 274  
 Db 3244 GAATAGTTAAAGATTTTGGGAATTCCTTTAATGTATAAAGAAATATTATATGTTT 3303  
 Qy 275 AsnIleAspTyrLeuAsnArgTyrMetTyr-----AlaAsnSerArgGlnIle 290  
 Db 3304 AATCGGGGAATAAAATTCATATATTAAACTAGTAGAAAGATTTCATCTAGGTGAAATA 3363  
 Qy 291 ValPheAsnThrArgArgAsnAsnAsn-----AspPheAsnGlu 303  
 Db 3364 TTAATACGTAGCAAAATATATACAGATTCCAATATATAAATATAGAAATTTATATATT 3423  
 Qy 304 GlyTyrLysIleIleLysArgIleArgGly-----AsnThrAsnAspThrArgVal 321  
 Db 3424 GGAGAAAAATTTATTATAAGAGAGAGAGTCAATTCATCTATATAATCATGATAGATT 3483  
 Qy 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLys----- 336  
 Db 3484 AGAAAGAAAGATTTATACATCTAGATTGTTACTTCAACATCAAGAGTGGAGCATATAT 3543  
 Qy 337 AlaTyrAsnLeuPheMetLysAsnGlu-----ThrMetTyrAlaAspAsn 351  
 Db 3544 GCCTATAAATTTTAAAGAACAGAGAGAAATTCGTTTTATCTATTATAAGTGAT--- 3600  
 Qy 352 HisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAsp 371  
 Db 3601 ---TCTAATGAATTTTATAAGACTATATAAATAAAGAA----- 3636  
 Qy 372 AsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePhe 391  
 Db 3637 -----TATGATGAACAGCCATCA-----TATAGTTCAGTTCGCTTTT 3675  
 Qy 392 LysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPhe 411  
 Db 3676 AAAAA-----GATGAAGAAAGTACATGATATAGGATTTGATTCATTCGTTTC 3729  
 Qy 412 ArgLeuGlyGly-----AspTyrTyrArgHisAsnTyrLeuVal----- 424  
 Db 3730 TAGGAATCTGAGCTTTTACGTAAAGATATAAGATTTATTTGTATAGTAATAGTAG 3789  
 Qy 425 ---ProThrValLysGlnGlyAsnTyrAlaSerLeuGluSerThrSerThrHisTrp 443  
 Db 3790 TTAAGAGAGTAAAGGAACCATATATAAGTCAAAATTTG-----GGATGAATG 3840  
 Qy 444 GlyPheValProValSerGlu 450  
 Db 3841 CAGTTTATCTCTAAAGATGAA 3861





[illegible]

QY 53 SerGluGluGlyAspValGlnLeuAsnProIlePheProPheAspPheLysLeuGlySer 72  
 Db 127 AACGTCGGTTCCGACGTCATCTTCAACGACATCGGTACGGTCAAGCTGAACCAAC 186  
 QY 73 SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrAsnSer 92  
 Db 187 TCC-----GAGAACTCCACATCACCAGCCGACCGTCCAGTTCGTCACGACATCC 240  
 QY 93 MetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpValSerAsn--- 111  
 Db 241 ATGTCGACAACTTCCATCACTTCTGGTCCGTACCCCAAGTACAAACACACAC 300  
 QY 112 -----LeuProGlyTyrThrIleLeuAspSerValIleAsnAsnSerGly 126  
 Db 301 ATCCAGACCTACTGCGAGACGAGTACACATCATCTCTGTATCAAGACGACATCCGGT 360  
 QY 127 TrpSerIleGlyIleIleSerAsnLeuValPheThrLeuLysGlnAsnGluAspSer 146  
 Db 361 TGGAAAGTCTCCATCAAGGAAACCGTATCATCTGACCCCTGATCGACGTCAACGCCAAG 420  
 QY 147 GluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaProGlyTyr---AsnLys 165  
 Db 421 TCCAAAGTCCATCTTCTCGAGTACTCCATCAAGACACATCTCCGACTACATCAACCAAG 480  
 QY 166 TrpPheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly 185  
 Db 481 TGGTTCCTCATCACCATCAACAGCAGCCGCTGGTGAACGCCAATCATACATCAACGGT 540  
 QY 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205  
 Db 541 TCCCTGAAGAAGTCGAGAGATCTCGACCTGGACCGTATCAACTCCCTCCCAACGACATC 600  
 QY 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAsnIle 225  
 Db 601 GACTTCAAGTGTGATGACTGTACCGACACCAACCAAGTTC----- 639  
 QY 226 AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn 245  
 Db 640 ---GTCTGGATCAAGGACTTCAACATCTTCGTCGGTGAAGTCAAGCCACCGAGGTCTCC 696  
 QY 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValIleValLysAspTyrTrpGlyAsnAspLeu 265  
 Db 697 TCCCTGTACTGTAGTCCAGTCTCCACCAACACACCTGAAGGACTTCTGGGGAACCCACTG 756  
 QY 266 ArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrIleAsnArgTyrMetTyrAla 285  
 Db 757 CGTACGACACCCAGTACTACCTGTTCAACCCAGGCTATGACGAAACATCTACATCAAGTAC 816  
 QY 286 AsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGlu----- 303  
 Db 817 TCTCCAGGCTCCATGGGTGAGACCGCCCTCGTACCAAC---TTCAACACGCCGCC 873  
 QY 304 -----GlyTyrLysIleIleIleLysArgIle----- 312  
 Db 874 ATCAACTACCAAGAACCTGTACCTGGGTCTCGTTCATCATCAAGAGCCCTCCCACTCC 933  
 QY 313 ArgGlyAsnThrAsnAspThrArgValArgGlyValAspIleLeuTyrPheAspMetThr 332  
 Db 934 CGTAACTCAACACACGACATCGTCCGTCAGGCTGACTACATCTACCTGACATC--- 990  
 QY 333 IleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHis 352  
 Db 991 -----GACAAACATC 999  
 QY 353 SerThrGluAspIleTyrAlaIleGlyLeuArgGluThrLysAspIleAsnAspAsn 372  
 Db 1000 TCCGACGAGTCC---TACCGTGTCTAGCTCTGTCACCTCCCAAGAGATCAGACCCAG 1056  
 QY 373 IleIlePheGlnIleGlnProMetAsnAsn-----ThrTyrTyrTyrAlaSerGlnIle 390  
 Db 1057 CTGTTC-----CTGGGCCCAATCAACGACGACCCCTACCTTCTACGACGCTCTCGACATC 1110

QY 391 -----PheLysSerAsnPheAsnGlyGluAsnIleSer----- 401  
 Db 1111 AAGAAGTACTAGAGAGACACCTACACTCTGAGTCTGCGGAGAGACACCAAG 1170  
 QY 402 -----GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArg 419  
 Db 1171 ACCTTCGGACTGTCGGTATCGTAACTGCTCAAGGACTACGCTTACGCTCTGGGACACC 1230  
 QY 420 His---AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSer 438  
 Db 1231 TAGGACAACTACTTC-----TGTATCTCCCGACTGGTACCTGCTGCTATCTCCGAGAAC 1284  
 QY 439 ThrSer-----ThrHisTrpGlyPheValProValSerGlu 450  
 Db 1285 ATCAACAAGCTGGTCTGGGATGTTACTGGCAGTTTCATCCACGTCGACGAG 1335

RESULT 11  
 US-09-350-756-7  
 ; Sequence 7, Application US/09350756  
 ; Patent No. US20020034521A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
 ; APPLICANT: John S. Lee  
 ; APPLICANT: Peter Pushko  
 ; APPLICANT: Michael D. Parker  
 ; APPLICANT: Jonathan F. Smith  
 ; APPLICANT: Mark T. Dertzbaugh  
 ; APPLICANT: Leonard Smith  
 ; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
 ; FILE REFERENCE: 003/124/SAP RIID 98-21  
 ; CURRENT APPLICATION NUMBER: US/09/350,756  
 ; CURRENT FILING DATE: 1999-07-09  
 ; EARLIER APPLICATION NUMBER: US 60/092,416  
 ; EARLIER FILING DATE: 1998-07-10  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
 ; SEQ ID NO 7  
 ; LENGTH: 2452  
 ; TYPE: DNA  
 ; ORGANISM: Clostridium botulinum  
 ; FEATURE:  
 ; US-09-350-756-7

Alignment Scores:  
 Pred. No.: 8,28e-55 Length: 2452  
 Score: 568.00 Matches: 149  
 Percent Similarity: 49.04% Conservative: 82  
 Best Local Similarity: 31.63% Mismatches: 184  
 Query Match: 23.76% Indels: 57  
 DB: 10 Gaps: 11

US-09-910-186A-10 (1-450) x US-09-350-756-7 (1-2452)

QY 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22  
 Db 1116 ATACCTTTTCAGCTTCCCAAAATACGATAGATATCAAGATTTATTATCTACTACTGAA 1175  
 QY 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42  
 Db 1176 TATATTAAAGATATTATTAATCTCTATATTGAATTAAGATATGAAAGTAAATCATTTA 1235  
 QY 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62  
 Db 1236 ATAGACTTATCTAGGTATGCATCAAAATAAATATTGGTAGTAAAGTAAATTTGATCCA 1295  
 QY 63 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79  
 Db 1296 ATAGATAAATCAAAATTCATTTATTATTTAGAAAGTAGT-----AAA 1340  
 QY 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99  
 Db 1341 ATTGAGGTAATTTTAAATAATGCTATTGTATATTAATAGTATGATGATAAATTTTACTACT 1400

Qy 100 SerPheThrPheArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116  
Db 1401 AGCTTTGGATAAGAAATCTCAAGTATTTACAGATTAAGTCAAAATTAATGAATATACA 1460  
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
Db 1461 ATAATAAATGTATGGAATAATTCAGGATGGAAGATCATCTTAATATGTTGAATA 1520  
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
Db 1521 ATCTGGACTTACAGGATACCTCAGGAATAAACAAGAGTAGTTTAAATAC----- 1574  
Qy 157 SerAsnAsnAlaProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMet 176  
Db 1575 AGTCA-AATGATTAATATATCAACAGATGATTTTGTACATCATCAATAATAGATA 1633  
Qy 177 GlyAsnMetLysIleValLeuAsnGlyLysLeuIleAspThrIleLysValLysGluLeu 196  
Db 1634 AATAACTCTAAATTTATATAATGGAAGATTAATAGATCAAAACCAATTTCAAAATTA 1693  
Qy 197 ThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGly 216  
Db 1694 GGTATATTCATGCTAGTATATATATTAATCTTTAAATAGATGTTGTAGATATACA--- 1750  
Qy 217 LeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAla 236  
Db 1751 -----CATAGATATATTTGGATAAAATATTTTAAATCTTTTGTAT 1789  
Qy 237 LysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnIleThrAsnVal 256  
Db 1790 AAGGAATTAATGAAAGAAATCAAGATTTATATGATTAATCAATCAATTCAGGTAT 1849  
Qy 257 ValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTrpMetValAsnIle 276  
Db 1850 TTAAGAAGCTTTGGGTGATTTATTAACAATATGATAAACCACTATATATTAATTA 1909  
Qy 277 AspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287  
Db 1910 TATGATCAATAATATGCGATGTAATAATAGTATAGGTATATAGAGTTATATGATCTT 1969  
Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAspPheAsnGlu 303  
Db 1970 AAAGGCGCTAGAGTAGGTAATGACTACAAACATTTATTTAAATCAAGTTGTATAGG 2029  
Qy 304 GlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGly 323  
Db 2030 GCGAATAATTTATATAAAATATGCTCTCGAAATTAAGATATATATGTTAGAAAT 2089  
Qy 324 GlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLys 343  
Db 2090 ATGATCGTGTATATATATGATAGTAGTAAATAAATAAAGATATAGGTTA----- 2140  
Qy 344 AsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArg 363  
Db 2141 -----GCTACTPAATGCATCAGCAGCGGTAGAAAATACTAAGTGA 2185  
Qy 364 GluGlnThrLysAspIleAsnAsnIleIlePheGlnIleGlnProMetAsnAsnThr 383  
Db 2186 TTGAAATACCTGATAGGAAATCTAAGTCAAGTAGTAGTATGAACTCAAAATATGAT 2245  
Qy 384 TyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIle 403  
Db 2246 CAAGGAATAACAATAAATGTC---AAATGAATTTACAAGATAATAATGGAATGATATA 2302  
Qy 404 CysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsnTyrLeu 423  
Db 2303 GCCTTATAGATTTTCATGCTTT-----AATAATATA 2335  
Qy 424 ValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440  
Db 2336 GCTAAAGTAGTAGCAAGTAATGTTGATATAAGACAAATAGAAAGATCTAGTAGGCTTGG 2395  
Qy 441 ---ThrHisTrpGlyPheValProValSerGlu 450

Db 2396 GGTGCTCATCGGAATTTTCTCTGTAGATGAT 2428  
RESULT 12  
US-10-051-952-3  
; Sequence 3, Application US/10051952  
; Patent No. US20020107199A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Patricia  
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
; FILE REFERENCE: 2933CIP  
; CURRENT APPLICATION NUMBER: US/10/051,952  
; CURRENT FILING DATE: 2002-01-17  
; PRIOR APPLICATION NUMBER: 09/730,237  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 3876  
; TYPE: DNA  
; ORGANISM: botulinum toxin  
US-10-051-952-3  
Alignment Scores:  
Pred. No.: 9,99e-54 Length: 3876  
Score: 561.00 Matches: 159  
Percent Similarity: 50.21% Conservative: 84  
Best Local Similarity: 32.85% Mismatches: 173  
Query Match: 23.46% Indels: 68  
Gaps: 21  
US-09-910-186a-10 (1-450) x US-10-051-952-3 (1-3876)  
Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuAspIleLeuAsnGlu 22  
Db 2506 ATGCGGTTTCATCTTCAATATATACCAATGATACAACTACTAATAGAAATGTTAATAA 2565  
Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerIleGluAsnArgLysAsnThrLeu 42  
Db 2566 TATAATAGCGAAATTTAAATAATATATCTTAAATTTAGATATAAGGATATAATTTA 2625  
Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluValAspValGlnLeuAsnPro 62  
Db 2626 ATAGATTTATCAGGATATGGGCAAGGTAGAGGTATATATGATGGAGTCGAGCTTAATGAT 2685  
Qy 63 IlePheProPheAspPheLysLeuGlySerSerGlyLysArgGlyLysValIleVal 82  
Db 2686 -----AAAAATCAATTTAAATTAAGTACTAGTCAGCA-----AATAGTAGATAGAGTG 2733  
Qy 83 ThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrp 102  
Db 2734 ACTCAAAATCAGATATACATATTTAATAGTGTCTCTCTGATTTAGCGTTAGCTTTGG 2793  
Qy 103 IleArgIleAsnLysTrp-----ValSerAsnLeu-----ProGlyTyrThr 116  
Db 2794 ATAAGAATACCTAAATAATAAGAATGATGTTATACAAATATATATCAATGAATATACA 2853  
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
Db 2854 ATATTAATTCGTGAAATAATATTCGGCTGGAATAATCATATTAGGGGTAATAGATA 2913  
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
Db 2914 ATATGGACTTTAATTTGATATAAATGGAATAAATCGGATATTTTGAATAATAACATA 2973  
Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
Db 2974 AGAAGAATATATCAGAGTATATAAATAGATGATGTTTGTACTACTATTACTATAAT--- 3030  
Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlnLysLeuIleAspThrIleLysValLysGlu 195  
Db 3031 TTGAATAACGCTAAATTTATATTAATGTTAAGCTAGTAATCAATCAATACAGATATAAAGAT 3090

```

FILE REFERENCE: 003/124/SAP RIID 98-21
CURRENT APPLICATION NUMBER: US/09/350.756
CURRENT FILING DATE: 1998-07-09
EARLIER APPLICATION NUMBER: US 60/092.416
EARLIER FILING DATE: 1998-07-10
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Apple Macintosh Microsoft Word 6.0
SEQ ID NO 1

```

```

; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-1

```

Alignment Scores:		
Prepd. NO.:	5.35e-54	Length: 1338
Score:	557.50	Matches: 144
Percent Similarity:	48.80%	Conservative: 80
Best Local Similarity:	31.37%	Mismatches: 180
Query Match:	23.32%	Indels: 55
DB:	10	Gaps: 11

US-09-910-186A-10 (1-450) x US-09-350-756-1 (1-1338)

16	LeuLysAspIleIleAsnGluTyrPheAsnAnsnIleAsnAspSerLysIleLeuSerLeu	39
QY		
18	CTGTGTCCTACCTTGAAATCATCAAGAACATCATCAATACCTCCATCCCGAACCTG	77
Db		
36	GlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlu	55
QY		
78	CGCTACGAATCCAATCACCTGTGCAGACTGTCTCGCTACGGTTCCAAAATCAACATCGGT	137
Db		
56	GlyAspValGlnLeuAsnProIlePheProPheAsp-----PheLysLeuGlySer	72
QY		
138	TCTAAGAATTAACTTCGATCGATCGCAGACAATCAGATCCAGCTGTTCAAATCTGGAATCT	197
Db		
73	SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrAsnSer	92
QY		
198	TCC-----AAATCGAAGTTATCTCGTAAGAATAGCTATCGTATACAGACTCT	242
Db		
93	MetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTyrPValSerAsnLeu	112
QY		
243	ATGTAGGAATAACTTCCACCCTCTCTCGATCCGTATCCCGAAATACTTCAACTCCATC	302
Db		
113	Pro-----GlyTyrThrIleIleAspSerValLysAsnAsnSerGlyTyrSerIle	129
QY		
303	TCTCTGAACAATGAATACACCATCATCAACTGCATGGAAAAACAATCTGGTGTGAAAGTA	362
Db		
130	GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer	149
QY		
363	TCTCTGGAACCTACGCTGGAATCATCTGGACTCTGCAGGACACTCAGAGAAATCAAAACCGGT	422
Db		
150	IleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr--AsnLysTyrPhePhe	168
QY		
423	GIYGATTTCAATACTCTCAGATGATCAACATCTCTGACTACATCAATCGTGGATCTTC	482
Db		
169	ValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeulle	188
QY		
483	GTTACCATCACCAACAATCTGCTGAAATACTCCAAATUTACATCAACGGCGCTGTATC	542
Db		
189	AspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGlu	208
QY		
543	GACCAGAACCGCATCTCCAATCTGGGTAACTCCAGCGTTCTAATAACATCATCTTCAAA	602
Db		
209	IleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrp	228
QY		
603	CTGGACGGTGTCTGTGACACT-----CACCCTCATCTCGG	638
Db		
229	IleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe	248
QY		
639	ATCAATACTTCATCTGTTTCGACAAGAACTTAACGAAAAAGAAATCAAAAGACTGTGAC	698
Db		

3850 CCTAAAGATGAA 3861

US-09-350-756-1  
; Sequence 1, Application US/09350756  
; Patent No. US20020034521A1

APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases

APPLICANT: JOHN S. LEE

; APPLICANT: Peter Pushko

APPLICANT: Michael D. Parker

APPLICANT: Jonathan F. Smith

; APPLICANT: Mark T. Dertzbaugh

APPLICANT: Leonard Smith  
TITLE OF INVENTION: Botulinum Neurotoxin Vaccine



QY 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspPheTyrAlaIleGlyLeu 362  
Db 3580 -----GCTACTATGATCATCACAGGCGGCTAGAAAAAATACTAAGT 3621  
QY 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382  
Db 3622 GCATTAGAAATACCTGATAGGAATCTAAGTCAAGTAGTAGTAATCAAGTCAAAAAAT 3681  
QY 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402  
Db 3682 GATCAAGGAATAACAATAAATGCG---AAATGAATTTACAGATAAATAATGGGAATGAT 3738  
QY 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrArgHisAsnTyr 422  
Db 3739 ATAGGCTTTATAGGAATTCATCAGTTT-----AATAAT 3771  
423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440  
3772 ATAGCTAACTAGTACAGATGATTTGGTATATAGACAAATAGAAAGATCTAGTAGGACT 3831  
441 -----ThrHisTyrGlyPheValProValSerGlu 450  
3832 TTGGTTGCTCATGGCAATTTATCTCTGTAGATGAT 3867  
BUILT 15  
US-09-350-756-9  
Sequence 9, Application US/09350756  
Patent No. US20020034521A1  
GENERAL INFORMATION:  
APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
APPLICANT: John S. Lee  
APPLICANT: Peter Pushko  
APPLICANT: Michael D. Parker  
APPLICANT: Jonathan F. Smith  
APPLICANT: Mark T. Dertzbaugh  
APPLICANT: Leonard Smith  
TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
FILE REFERENCE: 003/124/SAP RIID 98-21  
CURRENT APPLICATION NUMBER: US/09/350/756  
CURRENT FILING DATE: 1999-07-09  
EARLIER APPLICATION NUMBER: US 60/092,416  
EARLIER FILING DATE: 1998-07-10  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Apple Macintosh Microsoft Word 6.0  
SEQ ID NO 9  
LENGTH: 1327  
TYPE: DNA  
ORGANISM: Clostridium botulinum  
FEATURE:  
US-09-350-756-9  
Segment Scores:  
Seq. No.: 3 75e-53 Length: 1327  
Score: 550.00 Matches: 146  
Percent Similarity: 48.39% Conservative: 80  
Best Local Similarity: 31.26% Mismatches: 174  
Query Match: 23.00% Indels: 67  
DB: Gaps: 12  
US-09-910-186A-10 (1-450) x US-09-350-756-9 (1-1327)  
QY 16 LeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeu 35  
Db 10 TTATATCTACATTTACTGAATATATAGAAATATTATTAATCTCTATATGAAATTTA 69  
QY 36 GlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlu 55  
Db 70 AGATATGAAGTAATCATTTTANTAGCTATCTATGATGATGCATCAAAATAATATTGGT 129  
QY 56 GlyAspValGlnLeuAsnProIlePheProPheAsp-----PheLysLeuGlySer 72  
Db 130 AGTAAAGTAAATTTTGATCCCAATAGATAAAATCAAATTCATTTATTTATAGAAAGT 189

QY 73 SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrAsnSer 92  
Db 190 AGT-----AAATTCAGGTAAATTTTAAAAATGCTATTCGTATATAAATAGT 234  
QY 93 MetTyrGluSerPheSerIleSerPheTyrPheArgIleAsnLysTyrPheValSerAsnLeu 112  
Db 235 ATGTATGAAATTTTACTAGTCTTTGGTAAAGAAATCTCTAAGTATTTTAAACAGTATA 294  
QY 113 Pro-----GlyTyrThrIleLeuAspSerValLysAsnAsnSerGlyTyrSerIle 129  
Db 295 AGCTTAATTAATGAATATACATATAATTAATGTGTGGAATAATATTCAGGATGGAAGTA 354  
QY 130 GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGluSer 149  
Db 355 TCACCTAATTTATGTTGGAATAATCTGGACTTTACAGGATCTACAGGAATAAACAACAAGA 414  
QY 150 IleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr---AsnLysTyrPhePhe 168  
Db 415 GTAGTTTTTAATATACAGTCAATGATTAATATATACAGATATATATAACAGATGATTTT 474  
QY 169 ValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIle 188  
Db 475 GTAACTATCACAATAATAGATTAATAAATCTAAATTTATATAAATGGAAGATTAATA 534  
QY 189 AspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGlu 208  
Db 535 GATCAAAACCAATTTCAATTTAGTAAATTCATGCTAGTAAATATATATATGTTTAAA 594  
QY 209 IleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTyr 228  
Db 595 TTAGATGGTTGTAGAGATACA-----CATAGATATATTTGG 630  
QY 229 IleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe 248  
Db 631 ATAAATATTTTAAATCTTTTTCATAGGAATTAATGAAGAAATCAAGATTTATAT 690  
QY 249 AsnSerLeuGlnTyrThrAsnValLysAspTyrTyrPheLysAsnLeuArgTyrAsn 268  
Db 691 GATAATCAATCAAAATTCAGGTATTTTAAAGACTTTTGGGTGATTTTACAATATGAT 750  
QY 269 LysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer--- 287  
Db 751 AAACCATACTATATGTTAAATTTATATGATCCAAATAAATATGTCGATGAATAATGTA 810  
QY 288 -----ArgGlnIleValPheAsnThrArgArg 296  
Db 811 GSTATTAGAGTTATATGATCTTAAAGGCGCTAGAGTAGCGTAATGACTACA----- 864  
QY 297 AsnAsnAsnAspPhe-Asn-----GluGlyTyrLysIleI 308  
Db 865 -----AATGACTACAACATTTTAAATTCAGGTTTGTATAGGGGACAAATTTAT 918  
QY 308 eileLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTy 328  
Db 919 TATATAAAATATGCTTCGGAATAAAGATAATATGTTAGAAATAATGATCGTGATA 978  
QY 328 rPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTy 348  
Db 979 TATTAATGTAGTAGTTTAAATAAAGAAATATAGGTTA----- 1015  
QY 348 rAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAs 368  
Db 1016 -GCTACTATGCTACAGCGGCGCTAGAAAAAACTAAAGTCATTAAGTAATACCTGA 1074  
QY 368 pIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSe 388  
Db 1075 TGTAGGAATCTTAAGTCAAGTAGTAGTAATGAAGTCAAAAAATGATCAAGGAATAACAAA 1134  
QY 388 xGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyTh 408  
Db 1135 TAAATGC---AAATGAATTTACAGATAATTAATGGAATGATATAGGCTTTATAGGATT 1191  
QY 408 rTyrArgPheArgLeuGlyAspTyrTyrArgHisAsnTyrLeuValProThrVally 428

```

Db 1192 TCATCAGTTT-----|||   |||
QY 428 scInGlyAsnTyAlaSerLeuLeuGluSerThrSer-----|||
Db 1225 AACTAATTGGTATATAGACAAATAGAAAGATCTAGTAGGACTTTGGGTTGCTCATGGGA 1284
QY 444 yPheValProValSerGlu 450
Db 1285 ATTTATCTCTGATGATGAT 1303

```

Search completed: November 7, 2002, 20:06:56  
 Job time : 86 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 17:48:22 ; Search time 3151 Seconds  
(without alignments)  
3590.644 Million cell updates/sec

Title: US-09-910-186A-10  
Perfect score: 2391  
Sequence: 1 MTIFNFTSTNNLLKDII.....NYASLLESTHWGFPVPS 450

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Best-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

MODEL=frame+p2n.model -DEV=xlh  
-cgcn2\_1/USPTO.spool/US09910186/runat\_04112002.111618.20892/app.query.fasta.1.647  
-cgcn2\_Pending\_Patents\_NA\_Main -QFWT-fasap -SUFFIX=rrpm -MINMATCH=0.1 -LOOPCL=0  
-CGOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
USER=US09910186.ecgn.1.1.1396.grnat\_04112002.111618.20892 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_XMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THRSADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6 -FCAPEXT=7  
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Pending\_Patents\_NA\_Main:

1: /cgcn2\_6/ptodata/1/pna/pctus\_comb.seq.\*  
2: /cgcn2\_6/ptodata/1/pna/US06\_comb.seq.\*  
3: /cgcn2\_6/ptodata/1/pna/US07\_comb.seq.\*  
4: /cgcn2\_6/ptodata/1/pna/US080\_comb.seq.\*  
5: /cgcn2\_6/ptodata/1/pna/US081\_comb.seq.\*  
6: /cgcn2\_6/ptodata/1/pna/US082\_comb.seq.\*  
7: /cgcn2\_6/ptodata/1/pna/US083\_comb.seq.\*  
8: /cgcn2\_6/ptodata/1/pna/US084\_comb.seq.\*  
9: /cgcn2\_6/ptodata/1/pna/US085\_comb.seq.\*  
10: /cgcn2\_6/ptodata/1/pna/US086\_comb.seq.\*  
11: /cgcn2\_6/ptodata/1/pna/US087\_comb.seq.\*  
12: /cgcn2\_6/ptodata/1/pna/US088\_comb.seq.\*  
13: /cgcn2\_6/ptodata/1/pna/US089\_comb.seq.\*  
14: /cgcn2\_6/ptodata/1/pna/US090\_comb.seq.\*  
15: /cgcn2\_6/ptodata/1/pna/US091\_comb.seq.\*  
16: /cgcn2\_6/ptodata/1/pna/US092\_comb.seq.\*  
17: /cgcn2\_6/ptodata/1/pna/US093\_comb.seq.\*  
18: /cgcn2\_6/ptodata/1/pna/US094\_comb.seq.\*  
19: /cgcn2\_6/ptodata/1/pna/US095a\_comb.seq.\*  
20: /cgcn2\_6/ptodata/1/pna/US095b\_comb.seq.\*  
21: /cgcn2\_6/ptodata/1/pna/US095c\_comb.seq.\*  
22: /cgcn2\_6/ptodata/1/pna/US095d\_comb.seq.\*  
23: /cgcn2\_6/ptodata/1/pna/US096a\_comb.seq.\*  
24: /cgcn2\_6/ptodata/1/pna/US096b\_comb.seq.\*  
25: /cgcn2\_6/ptodata/1/pna/US096c\_comb.seq.\*  
26: /cgcn2\_6/ptodata/1/pna/US096d\_comb.seq.\*  
27: /cgcn2\_6/ptodata/1/pna/US096e\_comb.seq.\*  
28: /cgcn2\_6/ptodata/1/pna/US097a\_comb.seq.\*

29: /cgcn2\_6/ptodata/1/pna/US097b\_comb.seq.\*  
30: /cgcn2\_6/ptodata/1/pna/US097c\_comb.seq.\*  
31: /cgcn2\_6/ptodata/1/pna/US098a\_comb.seq.\*  
32: /cgcn2\_6/ptodata/1/pna/US098b\_comb.seq.\*  
33: /cgcn2\_6/ptodata/1/pna/US098c\_comb.seq.\*  
34: /cgcn2\_6/ptodata/1/pna/US099a\_comb.seq.\*  
35: /cgcn2\_6/ptodata/1/pna/US099b\_comb.seq.\*  
36: /cgcn2\_6/ptodata/1/pna/US099c\_comb.seq.\*  
37: /cgcn2\_6/ptodata/1/pna/US099d\_comb.seq.\*  
38: /cgcn2\_6/ptodata/1/pna/US100a\_comb.seq.\*  
39: /cgcn2\_6/ptodata/1/pna/US100b\_comb.seq.\*  
40: /cgcn2\_6/ptodata/1/pna/US101a\_comb.seq.\*  
41: /cgcn2\_6/ptodata/1/pna/US101b\_comb.seq.\*  
42: /cgcn2\_6/ptodata/1/pna/US102a\_comb.seq.\*  
43: /cgcn2\_6/ptodata/1/pna/US102b\_comb.seq.\*  
44: /cgcn2\_6/ptodata/1/pna/US6000\_comb.seq.\*  
45: /cgcn2\_6/ptodata/1/pna/US6001\_comb.seq.\*  
46: /cgcn2\_6/ptodata/1/pna/US6002\_comb.seq.\*  
47: /cgcn2\_6/ptodata/1/pna/US6003\_comb.seq.\*  
48: /cgcn2\_6/ptodata/1/pna/US6004\_comb.seq.\*  
49: /cgcn2\_6/ptodata/1/pna/US6005\_comb.seq.\*  
50: /cgcn2\_6/ptodata/1/pna/US6006\_comb.seq.\*  
51: /cgcn2\_6/ptodata/1/pna/US6007\_comb.seq.\*  
52: /cgcn2\_6/ptodata/1/pna/US6008\_comb.seq.\*  
53: /cgcn2\_6/ptodata/1/pna/US6009\_comb.seq.\*  
54: /cgcn2\_6/ptodata/1/pna/US6010\_comb.seq.\*  
55: /cgcn2\_6/ptodata/1/pna/US6011\_comb.seq.\*  
56: /cgcn2\_6/ptodata/1/pna/US6012\_comb.seq.\*  
57: /cgcn2\_6/ptodata/1/pna/US6013\_comb.seq.\*  
58: /cgcn2\_6/ptodata/1/pna/US6014\_comb.seq.\*  
59: /cgcn2\_6/ptodata/1/pna/US6015\_comb.seq.\*  
60: /cgcn2\_6/ptodata/1/pna/US6016\_comb.seq.\*  
61: /cgcn2\_6/ptodata/1/pna/US6017\_comb.seq.\*  
62: /cgcn2\_6/ptodata/1/pna/US6018\_comb.seq.\*  
63: /cgcn2\_6/ptodata/1/pna/US6019\_comb.seq.\*  
64: /cgcn2\_6/ptodata/1/pna/US6020\_comb.seq.\*  
65: /cgcn2\_6/ptodata/1/pna/US6021\_comb.seq.\*  
66: /cgcn2\_6/ptodata/1/pna/US6022\_comb.seq.\*  
67: /cgcn2\_6/ptodata/1/pna/US6023\_comb.seq.\*  
68: /cgcn2\_6/ptodata/1/pna/US6024\_comb.seq.\*  
69: /cgcn2\_6/ptodata/1/pna/US6025\_comb.seq.\*  
70: /cgcn2\_6/ptodata/1/pna/US6026\_comb.seq.\*  
71: /cgcn2\_6/ptodata/1/pna/US6027\_comb.seq.\*  
72: /cgcn2\_6/ptodata/1/pna/US6028\_comb.seq.\*  
73: /cgcn2\_6/ptodata/1/pna/US6029\_comb.seq.\*  
74: /cgcn2\_6/ptodata/1/pna/US6030\_comb.seq.\*  
75: /cgcn2\_6/ptodata/1/pna/US6031\_comb.seq.\*  
76: /cgcn2\_6/ptodata/1/pna/US6032\_comb.seq.\*  
77: /cgcn2\_6/ptodata/1/pna/US6033\_comb.seq.\*  
78: /cgcn2\_6/ptodata/1/pna/US6034\_comb.seq.\*  
79: /cgcn2\_6/ptodata/1/pna/US6035\_comb.seq.\*  
80: /cgcn2\_6/ptodata/1/pna/US6036\_comb.seq.\*  
81: /cgcn2\_6/ptodata/1/pna/US6037\_comb.seq.\*  
82: /cgcn2\_6/ptodata/1/pna/US6038\_comb.seq.\*  
83: /cgcn2\_6/ptodata/1/pna/US6039\_comb.seq.\*  
84: /cgcn2\_6/ptodata/1/pna/US6040\_comb.seq.\*  
85: /cgcn2\_6/ptodata/1/pna/US6041\_comb.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2391	100.0	1371	23	US-09-611-419A-9
2	2391	100.0	1371	34	US-09-910-186A-9
3	2386	99.8	3876	1	PCT-US97-15394-59
4	2386	99.8	3876	11	US-08-704-159-59
5	2386	99.8	3876	42	US-10-205-516-5
6	2386	99.8	3876	43	US-10-271-012-59

```
7 2386 99.8 3906 42 US-10-205-516-19
8 2386 99.8 3950 13 US-08-954-302-2
9 2381 99.6 1371 17 US-09-350-756-3
10 2364 98.9 3876 39 US-09-730-237-4
11 2364 98.9 3876 39 US-10-051-952-4
12 2319 97.0 1502 1 PCT-US97-15394-61
13 2319 97.0 1502 11 US-08-704-159-61
14 2319 97.0 1502 43 US-10-271-012-61
15 868.5 36.3 1374 23 US-09-611-419A-11
16 867.5 36.3 1374 23 PCT-US97-15394-65
17 867.5 36.3 3831 11 US-08-704-159-65
18 867.5 36.3 3831 42 US-10-205-516-7
19 867.5 36.3 3831 43 US-10-271-012-65
20 867.5 36.3 3861 42 US-10-205-516-21
21 867.5 36.3 3952 21 US-09-547-188-7
22 865.5 36.2 1374 34 US-09-910-186A-11
23 863.5 36.1 3831 29 US-09-730-237-5
24 863.5 36.1 3831 39 US-10-051-952-5
25 803.5 33.6 1469 1 PCT-US97-15394-67
26 803.5 33.6 1469 11 US-08-704-159-67
27 803.5 33.6 1469 43 US-10-271-012-67
28 604.5 25.3 3894 1 PCT-US97-15394-76
29 604.5 25.3 3894 11 US-08-704-159-76
30 604.5 25.3 3894 42 US-10-205-516-13
31 604.5 25.3 3894 43 US-10-271-012-76
32 604.5 25.3 3924 42 US-10-205-516-27
33 604.5 25.3 3937 21 US-09-547-188-13
34 602.5 25.2 3759 42 US-10-205-516-9
35 602.5 25.2 3789 42 US-10-205-516-23
36 595.5 24.9 3753 1 PCT-US97-15394-49
37 595.5 24.9 3753 11 US-08-704-159-49
38 595.5 24.9 3753 43 US-10-271-012-49
39 594.5 24.9 1400 34 US-09-910-186A-13
40 594.5 24.9 3759 1 PCT-US97-15394-51
41 594.5 24.9 3759 11 US-08-704-159-51
42 594.5 24.9 3759 43 US-10-271-012-51
43 594.5 24.9 4030 21 US-09-547-188-5
44 594.5 24.9 4030 21 US-09-547-188-9
45 593 24.8 3876 1 PCT-US97-15394-39
```

## ALIGNMENTS

## RESULT 1

Sequence 9, Application US/09611419A

GENERAL INFORMATION:

APPLICANT: Smith, Leonard A.

APPLICANT: Byrne, Michael P.

APPLICANT: Middlebrook, John L.

APPLICANT: Lapenotiere, Hugh

APPLICANT: Clayton, Michael A.

APPLICANT: Brown, Douglas R.

TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM

FILE REFERENCE: A3626 067252.0105

CURRENT FILING DATE: 2000-07-06

PRIOR FILING DATE: 2000-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

PRIOR FILING DATE: 1999-05-12

NUMBER OF SEQ ID NOS: 42

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on BONTA Hc
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)...(1359)
US-09-611-419A-9
```

## Alignment Scores:

```
Pred. No.: 4.07e-230 Length: 1371
Score: 2391.00 Matches: 450
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 23 Gaps: 0
```

US-09-910-186A-10 (1-450) x US-09-611-419A-9 (1-1371)

```
QY 1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIle 20
Db 10 ARGACATCCCATTCACATCTTCTCTACACCAAGCACTCCCTGTTGAAGGACATCATC 69
QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40
Db 70 AACGAGTAGTCTTCAACACATCAACAGCACTCCAGAGTCTCCCTGCGAAGACCTAAGA 129
QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60
Db 130 ACCTTGGTCGACACCTCCGTTACACGCCGAGGTCTCCGAGGAGGTCGACCTCCAGCTG 189
QY 61 AsnProIlePheProPheAspPheLysLeuGlySerGlyGluAspArgGlyLysVal 80
Db 190 AACCCAAATCTTCCCATTCGACTTCAAGCTGGTTCCTCCGCTGAGGACAGAGTAAGTC 249
QY 81 IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100
Db 250 ATCGTCACCCAGACCAACATCTGTCTACACTCCATGATAGGAGTCTCTTCCATCTCC 309
QY 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120
Db 310 TTCTGGATCAGAAATCAACAAGTGGTCTCCCACTTGGCAGGTACACCATCATCGACTCC 369
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
Db 370 GTCAAGAACAACTCCGTTGGTCTCATCGGTATCATCTCCAACTTCCCTGGTCTTCCAC 429
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAla 160
Db 430 AAGCAGAACGAGGAGTCCGAGCAGTCCCATCAACTTCTCTACGACATCTCCACACGCT 489
QY 161 ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180
Db 490 CTGTGGTACAAACAGTGGTCTTCTGTCACCTCCACCAACATGATGGGTAACTAAG 549
QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200
Db 550 ANCTACATCAACGGTAAGTGTATCGACACCATCAAGGTCAAGGAGTGTGACCGGTATCA 609
QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
Db 610 TTCTCCAGACCATCACCTTCGAGATCAACAGATCCCAAGACCCGCTGATCATCACTCC 669
QY 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheThrIlePheAlaLysGluLeuAsp 240
Db 670 GACTCCGACACATCAACATGTGGATCGGTGACTTCTACATCTTCCGCAAGGAGTGTGAC 729
QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnThrValThrAsnValLysAspTyr 260
Db 730 GGTAAGGACATCAACATCTCTTCAACTCTTGCAGTACACCAACGCTGTCAGGACTAC 789
```

```
QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsn 280
Db 790 TGGGGTAAGACCTGAGATACAAAGAGTACTACATGGTCAACATCGACTACTTGAAC 849
QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsp 300
Db 850 AGATACATGTAGCCCACTCCAGACAGATCTCTCTACACACAGACGTAACACACGAC 909
QY 301 PheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320
Db 910 TTCACGAGGGTTACAGATCATCATCAAGCGTATCAGAGGTAAACACACACACACAG 969
QY 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
Db 970 GTCAGAGGGTGTGACATCTGTCTACATTCACATGACTATCAACAACAGGCTTCAACCTG 1029
QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
Db 1030 TTCATGAGAACGAGACCATCATCAAGCGTATCAGAGGTAAACACACACACACACAG 1089
QY 361 GlyLeuArgGluGluThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380
Db 1090 GGTCTGCGTACGACACCAAGACATCAACACACATCATCTTCCAGATCCACCAATG 1149
QY 381 AsnAsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Db 1150 AACACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAAGACATC 1209
QY 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHis 420
Db 1210 TCGGATATCTGTTCCATCGGTACATCAAGATTCGGTCTGGTGGTGTGACTGGTACAGAC 1269
QY 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuGluSerThrSer 440
Db 1270 AACTACTGGTTCACACGTCTCAAGCAGGGTAACTACGGCTCTCTTCTGGAGTCCACITCC 1329
QY 441 ThrHisTyrGlyPheValProValSerGlu 450
Db 1330 ACCCACTGGGGATCGTCCAGTCTCCGAG 1359

RESULT 2
US-09-910-186A-9
Sequence 9, Application US/09910186A
GENERAL INFORMATION:
APPLICANT: U.S. Army Medical Research & Materiel Command
TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
FILE REFERENCE: A33626-A 067252.0107
CURRENT APPLICATION NUMBER: US/09/910.186A
PRIOR FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: PCT/US00/12890
PRIOR FILING DATE: 2000-05-12
PRIOR APPLICATION NUMBER: 09/611.419
PRIOR FILING DATE: 2000-07-05
PRIOR APPLICATION NUMBER: 60/133.865
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 60/133.866
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 60/133.867
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 60/133.868
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 60/133.869
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 60/133.873
PRIOR FILING DATE: 1999-05-12
PRIOR APPLICATION NUMBER: 08/123.975
PRIOR FILING DATE: 1993-09-21
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 1371
```

```
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
OTHER INFORMATION: sequence
NAME/KEY: CDS
LOCATION: (10)...(1359)
US-09-910-186A-9

Alignment Scores:
Pred. No.: 4,07e-230 Length: 1371
Score: 2391.00 Matches: 450
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 34 Gaps: 0

US-09-910-186A-10 (1-450) x US-09-910-186A-9 (1-1371)
QY 1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIle 20
Db 10 ATGACCATCCCATTCACATCTTCTCTACACCAACAACTCCCTGTTGAAGGACATCATC 69
QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40
Db 70 AACGAGTACTTCAACACATCAACGACTCCAAAGATCTCTCCCTGCAGAACCGTAAGAAC 129
QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60
Db 130 ACCTTGGTGGACACCTCCGGTTACAACGCCGAGGTCTCCGAGGAGGTGACGCCAGCTG 189
QY 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80
Db 190 AACCCAATCTTCCCATTCGACTTCAAGCTGGTTCCTCCGCTGAGGACAGAGGTAAAGTC 249
QY 81 IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100
Db 250 ATCGCACCCACAGACGAGACATCGTCTACAACTCCATGTACGAGTCTCTTCCATCTCC 309
QY 101 PheTyrIleArgIleAsnLysTyrPheValSerAsnLeuProGlyTyrThrIleLeuAspSer 120
Db 310 TTCCTGGATCAGATCAACAGTGGTCTCCAACTGCCAGGTACACCATCATCGATCC 369
QY 121 ValLysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
Db 370 GTCAAGACACACTCCGGTGTGCTTCCATCGGTATCATCTCCAATCTCTTGGTCTTCAACCTG 429
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
Db 430 AACGACAGGAGGACTCCGAGCGTCCATCAACTCTCTACGACATCTCCACACACGCT 489
QY 161 ProGlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180
Db 490 CCTGGTTACACAAAGTGTCTTCTCGTACCGCTCAACAAACATGATGGGTAAACATGAAG 549
QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200
Db 550 ATCTACATCAACGGTAAGTATCGACACCATCAAGGTCAAGGAGTTCACCGGTATCAAC 609
QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
Db 610 TTCTCCACAGCATCACCTTCGAGATCAACAAAGATCCACAGACACCGGTCTGATCACCTCC 669
QY 221 AspSerAspAsnIleAsnMetThrPheArgPheTyrIlePheAlaLysGluLeuAsp 240
Db 670 GACTCGCAACATCAACATGTGGATCGGTCTGACTTCTACATCTTCGCCAAGGAGTTCGAC 729
QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyr 260
Db 730 GGTAAAGGACATCAACATCTCTTCAACTCTTGCAGTACACCAACGCTCGTCAAGGACTAC 789
QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsn 280
```

Copied from 10910186 on 05-05-2004

Db 790 TGGGTAAACGACCTGAGATACAAACAGGAGTACTACTGCTCAACATCGACTACTTGAAC 849  
Qy 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsp 300  
Db 850 AGATACATGTACGCACTCCAGACAGATCGTCTTCAACACGAGAGTAAACACACGAC 909  
Qy 301 PheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320  
Db 910 TTCACAGGAGGTTACAAAGATCATCATCAACGATCATCAGAGGTAAACACACACACAGA 969  
Qy 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340  
Db 970 GTCAGAGTGTGACATCCCTGACTTCGACATGACTATCAACACAGGCTTACACCTG 1029  
Qy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
Db 1030 TTCATGAAGAACGAGACCATGTACGCGACACCACTCCACCGAGGACATCTACGCCATC 1089  
Qy 361 GlyLeuArgGluGlnThrLysAspIleAspAsnIleIlePheGlnIleGlnProMet 380  
Db 1090 GGTGTGAGTACAGACACGAGGACATCAACGACACATCATCTCCAGATCCAGCCATG 1149  
Qy 381 AsnAsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400  
Db 1150 AACACACTTACTACTACGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGAACATC 1209  
Qy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyClyAspThrTyrArgHis 420  
Db 1210 TCCGGTATCTTCCATCCGCTACGATCTCCGATTCGGTCTGGGTGAGTGGTACAGACAC 1269  
Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440  
Db 1270 AACTACTGTGTCCAACTGCAACGAGGTAACCTAGCCCTCCTTGTGGAGTCCACTTCC 1329  
Qy 441 ThrHisTrpGlyPheValProValSerGlu 450  
Db 1330 ACCACTGGGATTCGCTCCAGTCTCCGAG 1359

RESULT 3  
PCT-US97-15394-59  
Sequence 59, Application PC/TUS9715394  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
APPLICANT: Thalley, Bruce S.  
TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
TITLE OF INVENTION: Botulinum Neurotoxin  
NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/15394  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPND-02304  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3876 base pairs

TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3873  
PCT-US97-15394-59  
Alignment Scores:  
Pred. No.: 4,77e-229 Length: 3876  
Score: 2386.00 Matches: 449  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.79% Indels: 0  
DB: 1 Gaps: 0  
US-09-910-186a-10 (1-450) x PCT-US97-15394-59 (1-3876)  
Qy 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsn 21  
Db 2527 ACATACCCCTTAAATATTTTTCATATACTATAATATCTTTTAAAGATATATTAAT 2585  
Qy 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
Db 2587 GAATATTTCAATATATTAATGATTCAAAAATTTTTCAGCCTACAAAACAGAAAAATACT 2646  
Qy 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluClyAspValGlnLeuAsn 61  
Db 2647 TTAGTGATACATCAGATATATAATGCAAGTGAAGTGAAGGCGATGTTTCAGCTTAAT 2706  
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
Db 2707 CCAATATTTCCATTTGACTTTAAATTAGTATTAGTTCAGGGGAGGATAGAGTTAAAGTTATA 2766  
Qy 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
Db 2767 GTACCCAGAAATGAAATATGTATATAATTTATGTAAGTTTAGCAATTAGTTTT 2825  
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
Db 2827 TGGATTAGATAAATAAATGAGTAAATTTTACCCTGATATATATATATGATAGTGT 2886  
Qy 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
Db 2887 AAAATAACCTCAGGTTGGAGTATAGGTATATTAGTAATTTTATTAGTATTACTTAAAA 2946  
Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaPro 161  
Db 2947 CAAATCAAGATAGTGAACAAAGTATATAATTTTAGTATATATATCAATATGCTCT 3006  
Qy 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
Db 3007 GGATACAAATAAATGGTTTTTTTGTAACTGTACTAACAAATATGATGGGAATATGAAGATT 3066  
Qy 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysLysLeuThrGlyIleAsnPhe 201  
Db 3067 TATATAATGGAATAATTAATAGATATATAAAGTTAAAGAACTAACTGGAATTATTTT 3126  
Qy 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221  
Db 3127 AGCAAACTATACATTTGAATTAATAAATAAATTTCCAGATACCGGTTTGTATTCTTCAGAT 3186  
Qy 222 SerAspAsnIleAsnMetTrpIleArgAspPheThrIlePheAlaLysGluLeuAspGly 241  
Db 3187 TCTGATACATCAATATGTGGTAAGAGATTTTATATATTTCCTTAAGAAATATAGATGT 3246  
Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261  
Db 3247 AAAGATATTAATATATATTATTAATAGCTTCAATATATCTTAATCTTGAAGATATTGG 3306  
Qy 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281





Db 3487 CGAGGAGGAGATATTATTTATTTGATATGCAATTAATAACAAGCATATAATTGTTT 3546  
QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361  
Db 3547 ATGAAGAATGAACATGATGATGCAGATATCATAGTCTGAAGATATATATGCTTAGGT 3606  
QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
Db 3607 TTAAGAGAACAACAAGGATATAATGATATATATTTTCAATACAAACCAATGAAT 3666  
QY 382 AsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
Db 3667 AATACATATATTAACGATCTCAATATTTAAATCAATTTTATGAGAAATATTTCT 3726  
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsn 421  
Db 3727 GCAATATGTTCAATAGTACTTATCGTTTATAGACTTGGAGGTGATGTTGATAGACAAAT 3786  
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441  
Db 3787 TATTTGGTCCCTACTGTGAGGCAAGAAATATATGCTTCATTATTAGATCAACATCAACT 3846  
QY 442 HisTyrGlyPheValProValSerGlu 450  
Db 3847 CATTTGGGTTTGTACCTGTAAGTGAA 3873

RESULT 6

US-10-271-012-59  
Sequence 59, Application US/10271012  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
Thalley, Bruce S.  
TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
Borulinum Neurotoxin  
NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Medien & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/271,012  
FILING DATE: 15-Oct-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/704,159  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPED-02304  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
TYPE: nucleic acid  
LENGTH: 3876 base pairs  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3873  
SEQUENCE DESCRIPTION: SEQ ID NO: 59:

US-10-271-012-59  
Alignment Scores:  
Pred. No.: 4,77e-229 Length: 3876  
Score: 2386.00 Matches: 449  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.79% Indels: 0  
DB: 43 Gaps: 0  
US-09-910-186a-10 (1-450) x US-10-271-012-59 (1-3876)  
QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
Db 2527 ACAATACCTTTAAATATTTTTCATATACCTAATAATCTTTTAAAGATATATTAAT 2586  
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAspThr 41  
Db 2587 GAATATTTCAATAATTAATGATTCAAAAATTTTGAGCTTACAAAACAGAAAAAATACT 2646  
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
Db 2647 TTAGTGATACATCAGGATATAATGCAGAAGTGAAGGAGGCGATGTTTCAGCTTAAT 2706  
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
Db 2707 CCAATATTTCCATTTGACTTTAAATTAGTAGTTAGTTCAGGGAGGATAGAGTTAAAGTTATA 2766  
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
Db 2767 GTAACCCAGATGAATAATTTGTATATAATCTATGTAAGAAAGTTTAGCATTAGTTT 2826  
QY 102 TrpIleArgIleAsnLysTyrValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
Db 2827 TGGATTAGAAATAAATAGGTGAAGTAATTTTACCTGGATATACTATAATATTGATAGTTT 2886  
QY 122 LysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
Db 2887 AAAAAATACTCAGGTGGAGTATAGGTATTTATAGTAATTTTGTAGTATTTTACTTTAAA 2946  
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaPro 161  
Db 2947 CAAATGAAGATAGTGACAAAGATATAATTTAGTTATGATATATCAATAATGCTCCT 3006  
QY 162 GlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
Db 3007 GGATACATATAATAGTTTGTAACTCTTACTACAAATATGATGGAAATATGAAAT 3066  
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
Db 3067 TATATAATGGAAATTTATAGATATCTATAAAGTTAAAGAACTAACTGGAAATTAATTT 3126  
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221  
Db 3127 AGCAAACTATACATTTGAAATAAATAAATTCAGATACCGGTTTGTACTTACAT 3186  
QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241  
Db 3187 TCGATACATCAATATGCGATAGACAGATTTTATATATTTGCTAAGAATTAGATGGT 3246  
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261  
Db 3247 AAAGATATTAATATATTTAATAGCTGCAATATACATAATGTTTAAAGATTTATTGG 3306  
QY 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281  
Db 3307 GGAATGATTTAGATATAATAAGAATATTATATGTTTATATAGATTATTTAAATAGA 3366  
QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
Db 3367 TATATGTATCGGAACCTCAGCAAAATTTGTTTTTAATACACGTAGAAATAATAATGACTTC 3426  
QY 302 AsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321

```

Db 3427 AATGAAGGATATAAAATTAATAAAAGAAATCAGAGGAATACAAATGATCTAGAGTA 3486
QY 322 AAGGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341
Db 3487 CGAGGAGGAGATATTATATATTTGATGATGACAAATTAATAACAAGCATATAATTTGTTT 3546
QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361
Db 3547 ATCAAGAATGAACATCATGATGACATTAATCATAGTACTGAAGATATATATGCTATAGGT 3606
QY 362 LeuArgGluGluThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
Db 3607 TTAAGAGAAACAACAGCATATAATGATATATATATATTTCAATACACCAATGAAT 3666
QY 382 AsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401
Db 3667 ATACTTATATATAGCATCTCAATATATTTAAATCAATTTTAAATGAGAGAAATATTTCT 3726
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsn 421
Db 3727 GGAATATGTTCAATAGTACTATCGTTTATAGACTTGGAGGTGATTGCTATAGACACAAT 3786
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
Db 3787 TATTGGTCCCTACTGTGAGCAAGAAATTAATGCTTCATTTATAGATCAATCAACT 3846
QY 442 HisTrpGlyPheValProValSerGlu 450
Db 3847 CATTTGGGGTTTGTACCTGTAAGTCAA 3873

SUBMIT 7
10-205-516-19
Sequence 19, Application US/10205516
GENERAL INFORMATION:
APPLICANT: Zhong, Jun
TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
FILE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
FILE REFERENCE: J2btxl
CURRENT APPLICATION NUMBER: US/10/205,516
CURRENT FILING DATE: 2002-07-25
NUMBER OF SEQ ID NOS: 28
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 19
LENGTH: 3906
TYPE: DNA
ORGANISM: Clostridium botulinum
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(3906)
NAME/KEY: misc_signal
LOCATION: (1347)..(1358)
OTHER INFORMATION: factor Xa site
FEATURE:
NAME/KEY: misc_signal
LOCATION: (3886)..(3903)
OTHER INFORMATION: 6-histidine tag
US-10-205-516-19

Alignment Scores:
Pred. No.: 4 82e-229 Length: 3906
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 42 Gaps: 0

US-09-910-186a-10 (1-450) x US-10-205-516-19 (1-3906)

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 2539 ACAATACCCCTTAAATATTTTTCATATACATAATATCTTTTATTTAAAAGATATAATTAAT 2598

```

Copied from 10910186 on 05-05-2004

```

QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2599 GAATATTTCAATATATATTAATGATTTCAAAATTTTGACCTACAAACACAGAAAATAACT 2658
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 2659 TTAGTGGATACATCAGGATATAATGCGAAGTGAGTGAAGAAGCGGATGTGACGTTAAT 2718
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 2719 CCAATATTTCCATTTGACTTTAAATTTAGTTTCAGGGAGGATAGAGTAAGTTATA 2778
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 2779 GTAACCCAGATGAATATGTTATATATCTATGTAATGAAAGTATTTAGCATTAGTTT 2838
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121
Db 2839 TCGATTAGAAATAAATAAATGGGTAAATTTACCTGGATATACTATAATTTGATAGTGT 2898
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 2899 AAAATAACTCAGGTTCAGTATAGTATATTTAGTATTTTACTTTTAAAT 2958
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaPro 161
Db 2959 CAAAATGAAGTAGTAGCAAAAGTATAAATTTTATGTTATGATATATCAATAATGCTCT 3018
QY 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3019 GGATACAATAATGTTTGTACTGTACTAACAATATGATGGGAATATGAAGATT 3078
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 3079 TATATAATGGAAATTAATAGACTATAAAGTTTAAAGACTTAAGTGAATTAATTT 3138
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 3139 ACRAAATCAATACATTTGAATAAATAAATTCAGATACCGGTTGATCTACTCAGAT 3198
QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241
Db 3199 TCTGATAACATCAATATGTTGATAGAGATTTTATATATTTGCTAAGAAATAGATGGT 3258
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261
Db 3259 AAGATATTTAATATATTTTAAATAGCTTGCATATATCTAATGTTGTAAGATTTATGG 3318
QY 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281
Db 3319 GGAAATGATTTAAGATATAATAAAGAAATATTATATGTTTATATATAGATTATTTAATAFAGA 3378
QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301
Db 3379 TATATGATGCGAAGTCAAGCAATTTGTTTAAATACACGACGATATAATATGACTTC 3438
QY 302 AsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321
Db 3439 AATGAAGGATATAAATAATTAATAAAGAAATCAGAGAAATAGATAATGATAGAGTA 3498
QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341
Db 3499 CGAGGAGGAGATATTTTATTTTATGATGACAAATTAATAACAAGCATATAATTTGTTT 3558
QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361
Db 3559 ATGAAGATGAACATCTATGTCAGATAATCATAGTACTGAGGATATATATGCTATAGGT 3618
QY 362 LeuArgGluGluThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
Db 3619 TTAAGAGAAACAACAAGGATATAAATGATAATATTTATTTCAATACCAATCAATGAAT 3678

```



QY 382 AsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
 DB 3679 AATACTATTATAGCGATCTCAAAATATTTAAATCAAAATTTTAAATGGAGAAATATTCT 3738  
 QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrArgHisAsn 421  
 DB 3739 GGAATATGTTCAATAGGTACATATCGTTTACAGCTTGAGGTGATTGGTATAGACACAAT 3798  
 QY 422 TyrLeuValProThrValIleGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441  
 DB 3799 TATTTGGTCCCTACTGTGAAGCAAGAAATTTATGCTTCATTATTAGAAATCAATCAACT 3858  
 QY 442 HisTrpGlyPheValProValSerGlu 450  
 DB 3859 CATGGGGTTTGTACCTGTAGTGA 3885

## RESULT 8

US-08-954-302-2  
 Sequence 2, Application US/08954302  
 GENERAL INFORMATION:  
 APPLICANT: Lance Simpson, Nikita Kiyatkin,  
 APPLICANT: Andrew Makymowych  
 TITLE OF INVENTION: Compositions and Methods for Systemic  
 Title of Invention: Delivery of Oral Vaccines and Therapeutic Agents  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: Jane Massey Licata, Esq.  
 STREET: 66 E. Main Street  
 CITY: Marlton  
 STATE: NJ USA  
 COUNTRY: USA  
 ZIP: 08053  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
 COMPUTER: IBM 486  
 OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
 SOFTWARE: WORDPERFECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/954,302  
 FILING DATE: herewith  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Jane Massey Licata  
 REGISTRATION NUMBER: 32,257  
 REFERENCE/DOCKET NUMBER: JEFF-0164  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (609) 779-2400  
 TELEFAX: (609) 810-1454  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3950  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: SINGLE  
 TOPOLOGY: LINEAR  
 ANTI-SENSE: NO  
 US-08-954-302-2

Alignment Scores:  
 Pred. No.: 4,896-229 Length: 3950  
 Score: 2386.00 Matches: 449  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.79% Indels: 0  
 DB: 13 Gaps: 0

US-09-910-186a-10 (1-450) x US-08-954-302-2 (1-3950)

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsn 21  
 DB 2558 ACAATACCCCTTTAAATATTTTTCATATCTATATAATCTTTTAAAGATATATTAAT 2617

QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
 DB 2618 GAATATTTCAATAATATTAATGATTTCAAAATTTTGAGCCTACAAAACAGAAAAATACT 2677  
 QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
 DB 2678 TTAGTGGATACATCAGGATATATGTCAGAACTGAGTGAAGAGCGGATGTTCAGCTTAAT 2737  
 QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
 DB 2738 CCAATATTTCCATTTGACTTTAAATTTAGTGTAGTTAGTTCAGGGGAGGATAGAGTAAAGTTATA 2797  
 QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
 DB 2798 GTAACCCAGATGAATAATTTGTATATAATCTATGATGAAGATTTTAGCATTTAGTTT 2857  
 QY 102 TyrIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
 DB 2858 TGGATTAGAATAAATAAATGGTAAATTTACCTGGATATACATAATTAATGATAGTGT 2917  
 QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
 DB 2918 AAAATAAATCAGGTGGAGTATAGGTATTTAGTAAATTTTTTAGTATTACTTTAAAA 2977  
 QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaPro 161  
 DB 2978 CAAATGAAGATAGTGAACAAAGATATAATTTTAGTTATGATATATCAATAATGCTCT 3037  
 QY 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
 DB 3038 GGATACAAATAAATGGTGTGTACTACTAACTATGATGGAATATGAAGATT 3097  
 QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
 DB 3098 TATATAATGGAATAATTAATAGATATAAAGTAAAGAACTAACTGGAATTAATTT 3157  
 QY 202 SerLysThrIleThrPheGluIleAsnLysIleProaspThrGlyLeuIleThrSerAsp 221  
 DB 3158 AGCAAACTATTAACATTTGAAATTAATAAATTCAGATACCGGTTTCATTACTTCAGAT 3217  
 QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241  
 DB 3218 TCTGATACATCAATATCTGGATAGAGATTTTATATTTGCTAAGAATTAGATGTT 3277  
 QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrp 261  
 DB 3278 AAAGATATTAATATATTTAAATAGCTTGAATATATACTAATGTGTAAAGATATTATGG 3337  
 QY 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281  
 DB 3338 GGAATGATTTAAGATATATAAAGATATTTATATGTTAATATAGATTATTAAATAGA 3397  
 QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
 DB 3398 TATATGATGCGAACTCAGCAAAATTTGTTTAAATACACGTAGAAATAATAATGACTTC 3457  
 QY 302 AsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
 DB 3458 AATGAAGGATATAAATAATATAAAGAAATCAGAGAAATACAAATGATACATAGAGTA 3517  
 QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
 DB 3518 CGAGGAGGAGATATTTTATTTTATGATGACAATTAATAACAAGCATATAATTTGTTT 3577  
 QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361  
 DB 3578 ATGAAGATGAAGAACTATGTCAGATTAATCATAGTACTGAGGATATATATGCTTATAGGT 3637  
 QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
 DB 3638 TTAAGAGACACAAAGGATATAATGATAATATATATTTTCAATACAAATACACCAATGAAT 3697

QY 382 AsnThrTyrTyrValSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
DB 3698 AATACTATTATTAAGGATCTCAATATTTTAAATCAAAATTTTAAATGGAGAAATATTTCT 3757  
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsn 421  
DB 3758 GGAATATGTTCAATAGGTACTATCTGTTTAGACTTGGAGGTGATTGGTATAGACACAAT 3817  
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerIleLeuGluSerThrSerThr 441  
DB 3818 TATTGGTGCCCTACTGTGACAGCAAGAAATATGCTTCATTATTAATCAACATCAACT 3877  
QY 442 HisTrpGlyPheValProValSerGlu 450  
DB 3878 CATGGGGTTTGTACCTGTAAGTGAA 3904

RESULT 9  
US-09-350-756-3  
Sequence 3, Application US/09350756  
GENERAL INFORMATION:  
APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
APPLICANT: John S. Lee  
APPLICANT: Peter Pushko  
APPLICANT: Michael D. Parker  
APPLICANT: Jonathan F. Smith  
APPLICANT: Mark T. Dertzbaugh  
APPLICANT: Leonard Smith  
TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
FILE REFERENCE: 003/124/SAP RIID 98-21  
CURRENT APPLICATION NUMBER: US/09/350,756  
CURRENT FILING DATE: 1999-07-09  
EARLIER APPLICATION NUMBER: US 60/092,416  
EARLIER FILING DATE: 1998-07-10  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Apple Macintosh Microsoft Word 6.0  
SEQ ID NO 3  
LENGTH: 1371  
TYPE: DNA  
ORGANISM: Artificial  
FEATURE:  
OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"  
US-09-350-756-3

Alignment Scores:  
Seq. No.: 4.13e-229 Length: 1371  
Score: 2381.00 Matches: 448  
Percent Similarity: 99.78% Conservative: 1  
Best Local Similarity: 99.56% Mismatches: 1  
Query Match: 99.58% Indels: 0  
Gaps: 17

US-09-910-186a-10 (1-450) x US-09-350-756-3 (1-1371)  
1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerIleLeuLysAspIleIle 20  
10 ATCCACATCCCATTCACATCTTCCTACACCAACACACCTCCCTGTTGAAGGACATCATC 69  
21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40  
70 AAGGAGTACTTCAACACATCAACAGCTCCCAAGATCCTGCTCCCTGCAGAACCGTAAAGAAC 129  
41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60  
130 ACCTTGCTGCACCTCCGGTTACACCCGAGGTCCTCCGAGGAGGTGACGTCCACCTG 189  
61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80  
190 AACCCCAATCTCCCATCGACTTCAAGCTGGTCTCCCTCCGCTGAGGACAGAGGTAAAGTC 249  
81 IleValThrGluAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100  
250 ATCGTCACCCAGACAGAGACATCTCTACAACTCCATGTACGAGTCTCTTCCATCTCC 309

QY 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120  
DB 310 TTCGGATCAGATCAACAAGTGGTCTCCAACTTCCAGGTTACACCATCATCGACTCC 369  
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140  
DB 370 GTCAGAACAACTCCGGTTGGTTCATCGGTATCATCTCCACCTTCTTGGTCTTCAACCTG 429  
QY 141 LysGlnAsnGluAspSerGluInSerIleAsnPheSerTyrAspIleSerAsnAla 160  
DB 430 AAGCAGAACAGGACTCCGAGAGTCCATCAACTTCTCTACGACATCTCCAAACAGCT 489  
QY 161 ProGlyTyrAsnLysTrpPheValThrValThrAsnAsnMetMetGlyAsnMetLys 180  
DB 490 CCTGGTTACAACAAGTGGTCTTCTCGTCACGTCACCAACAACATGATGGTAACTGAAG 549  
QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysLeuThrGlyIleAsn 200  
DB 550 ATCTCATCAACGGTAAGCTGATCGACACCATCAAGGTCAAGAGTTCACCGGTATCAAC 609  
QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyIleIleThrSer 220  
DB 610 TTCCTCAAGACCATCACTCTCGAGATCAACAAGTCCACACACCGGTCTGATCACTCC 669  
QY 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240  
DB 670 GACTCCGACATATCAACATGTGGATCCGTGACTTCTACATCTTCCCAAGGAGTGGAC 729  
QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyr 260  
DB 730 GGTAAAGACATCAACATCTCTCACTCTGACGTACACCAACCGTCTCAAGGACTAC 789  
QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrIleuAsn 280  
DB 790 TGGGTAAAGACATCAACATGTGGATCCGTGACTTCTACATCTTCCCAAGGAGTGGAC 849  
QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAsp 300  
DB 850 AGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACACGAGGTAAACAACAGAC 909  
QY 301 PheAsnGluGlyTyrLysIleIleIleIleIleIleIleIleIleIleIleIleIleIle 320  
DB 910 TTCACAGAGGTTACAGATCATCATCAAGGTATCAGAGGTAAACCAACAGCACCA 969  
QY 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340  
DB 970 GTCAGAGGTGGTGGTCTCTGACTTCGACATGACTATCAACAACAGGCTTACAACCTG 1029  
QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
DB 1030 TTCATGAAGACAGAGACATGTACCGCGAACCACTCCACCGAGGACATCTACGCCATC 1089  
QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPromet 380  
DB 1090 GGTCTGCGTGGAGACACCAAGGACATCAACGACAACTATCTTCCAGATCCAGCCAATG 1149  
QY 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400  
DB 1150 AACAAACATCTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAACATC 1209  
QY 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHis 420  
DB 1210 TCCGATCTCTTCCATCGGTACCTACAGATTCGCTGGTGGTGGTGGTGGTGGTGGTGG 1269  
QY 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440  
DB 1270 AACTACTTGGTTCCACTGTCAAGCAGGTAACTACGCCCTCTTGTGGAGTCCACTTCC 1329  
QY 441 ThrHisTrpGlyPheValProValSerGlu 450  
DB 1330 ACCCACTGGGATTCGTCCTCCAGTCTCCGAG 1359  
RESULT 10

```

US-09-730-237-4
; Sequence 4, Application US/09730237
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Treating Hyperhidrosis
; FILE REFERENCE: 2933
; CURRENT APPLICATION NUMBER: US/09/730,237
; CURRENT FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3876
; TYPE: DNA
; ORGANISM: botulinum toxin
US-09-730-237-4

Alignment Scores:
Pred. No.: 7,83e-227 Length: 3876
Score: 2364.00 Matches: 447
Percent Similarity: 99.55% Conservative: 0
Best Local Similarity: 99.55% Mismatches: 2
Query Match: 98.87% Indels: 0
DB: 29 Gaps: 0

US-09-910-186a-10 (1-450) x US-09-730-237-4 (1-3876)
2 ThrilleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
2527 ACAATACCCCTTAATATTTTTCATATACATAAATCTTTATTAAGAATATAATTAAT 2586
22 GluTyPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
2587 GAATATTTCAATATTAATTAATGATCAAAATTTTGAGCCCTACAAAACAGAAAATACT 2646
42 LeuValAspThrSerGlyTyAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
2647 TTATGGGATACATCAGATATAATGACAGAGCTGAGTGAAGAGCGATGTCAGCTTAAT 2706
62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
2707 CCAATATTTCCATTTGACTTTAAATTAGTATGTTAGTTCAGGGGAGGATAGAGTAAAGTTATA 2766
82 ValThrGlnAsnGluAsnIleValTyAsnSerMetTyThrGluSerPheSerIleSerPhe 101
2767 GTAACCCAGATGAATAATGTAATTAATTTCTATGTAAGAACTTTTAGCATAGTTT 2826
102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyThrIleIleAspSerVal 121
2827 TGGATTAGAATAAATAATGGTAAAGTAAATTTACCTGGATATACATAATTAATGATGTT 2886
122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
2887 AAAATAACTCAGGTGGAGTATAGGTATTAATTAATTTTATGATTTTACCTTTAAAA 2946
142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyAspIleSerAsnAlaPro 161
2947 CAAATGAAGATAGTGAACAAAGTATAAATTTAGTTATGATATACAAATAATGCTCCT 3006
162 GlyTyAsnLysTrpPheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
3007 GGATAGATAAATAGTTTGTGTAAGTCTTACTACAAATATGATGGGAATATGAAGATT 3066
182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
3067 TATATAAATGGAATTAATAGATACATAAAGTTAAAGAACTAACTGGAATTAATTT 3126
202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
3127 AGCAAACTATACATTTTGAATAAATAAATTCAGATACCGGTTTGATTTACTTCAGAT 3186
222 SerAspAsnIleAsnMetTrpIleArgAspPheTyThrIlePheAlaLysGluLeuAspGly 241
3187 TCTGATACATCAATATGCGGATAGAGATTTTATATATTTTGTATAGAAATTAAGATGCT 3246

```

QY 2 ThrileProPheAsnIlePheSerTyrThrAsnAspSerLeuLeuLysAspIleLeu 21  
DB 2527 ACAATACCCCTTTAATATTTTTCATATACATATATATATATATATATATATAT 2586  
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
DB 2587 GAATATTTCAAT 2646  
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyValGlnLeuAsn 61  
DB 2647 TTAGTGGATACATCAGGATATATATGAGAGAGTGGAGAGGCGATGTTGAGCTTAT 2706  
QY 62 ProIlePhePropheAspPheLysLeuGlySerSerGlyIleAspArgGlyLysValIle 81  
DB 2707 CCAATATTTCCATTGCTTTAAATAGTAGTTTCAGGGGAGGATAGAGGTAAGTTATA 2766  
82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrClnSerPheSerIleSerPhe 101  
2767 GTACCCAGAGTAAAT 2826  
102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
2827 TGGATTAGATAAATAAATGGTAAAGTAATTTACCTGGATATATATATATATATAT 2886  
122 LysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPhelValPheThrLeuLys 141  
2887 AAAATAACTCAGGTTGGAGTAGGATATATATATATATATATATATATATATAT 2946  
142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161  
2947 CAAATGAAGATAGTGAACAAAGTATAAATTTAGTTATGATATATCAAAATATGCTCT 3006  
162 GlyTyrAsnLysTrpPheValThrValThrAsnAsnMetMetIleAsnMetLysIle 181  
3007 GGATAGATAAATGGTTTTTGTGTAAGTGTACTAAATATGATGGAATATGAGATT 3066  
182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
3067 TATATAATGCAAAATATATAGATATATATAAAGTAAAGAACTAACTGGAATATTTT 3126  
202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221  
3127 AGCAAACTATAAATATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3186  
222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241  
3187 TCTGTATACATCAATATGTGATAGAGATTTTATATATTTCTTAAGAAATAGATGT 3246  
242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrp 261  
3247 AAAGAT 3306  
262 GlyAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrIleAsnArg 281  
3307 GGAAATGATTTAAGATATATATATATATATATATATATATATATATATATAT 3366  
282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
3367 TATATGTATGCAACTCAGCAAAATTTGTTTAAATACACGTAGAAATAATATGACTTC 3426  
302 AsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
3427 AATGAAGATATATAAATATATAAAGAAATCAGAGGAAATCAAAATGATAGATGA 3486  
322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
3487 CGAGGAGGAGATATTTATATATTTGATGACATTAATAACAAGCATATATTTGTT 3546  
342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361  
3547 ATGAAGATGAACATGATGATGAGATATATATATATATATATATATATATATAT 3606

QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
DB 3607 TTAGAGACAACAACAAGGATATAAATGATATATATATATATATATATATATAT 3666  
QY 382 AsnThrTyrTyrThrAlaSerGlnIlePheLysSerAsnPhelAsnGlyGluAsnIleSer 401  
DB 3667 AATATCTTATTTAGGATCTCAATATATTTAAATCAAAATTTTAAATGAGAAATATTTCT 3726  
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyLysAspTyrArgHisAsn 421  
DB 3727 GGATATGTTCAATAGTACTTATCTTTTAGACTTGGAGGTGATTTGATATAGACAAAT 3786  
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuSerThrSerThr 441  
DB 3787 TATTTGTGCTACTGTGAAGCAAGAAATATGCTTCATTTATTAGAAATCAACATCACT 3846  
QY 442 HisTrpGlyPheValProValSerGlu 450  
DB 3847 CATGGGGTTTTGTACCTGTAACTGAA 3873

RESULT 12  
PCT-US97-15394-61  
; Sequence 61, Application PC/TUS9715394  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; TITLE OF INVENTION: Botulinum Neurotoxin  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medien & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentID Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/15394  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 61:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1502 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 108..1493  
PCT-US97-15394-61

Alignment Scores:  
Pred. No.: 8,09e-223 Length: 1502  
Score: 2319.00 Matches: 439  
Percent Similarity: 98.66% Conservative: 3  
Best Local Similarity: 97.99% Mismatches: 5  
Query Match: 96.99% Indels: 1  
DB: 1 Gaps: 0

US-09-910-186A-10 (1-450) x PCT-US97-15394-61 (1-1502)

QY 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleLeuAsnGluTy 23  
| | | | | : : : : :  
Db 152 CCATATCGAGGTCGTCATATAGGCTAGCATGCTTTATTAAGATATAATATGATA 211  
| | | | | : : : : :  
QY 23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43  
| | | | | : : : : :  
Db 212 TTTCATATATATATGATTCAAAATTTTGAGCCTACAAACAGAGAAAATACCTTAGT 271  
| | | | | : : : : :  
QY 43 lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIl 63  
| | | | | : : : : :  
Db 272 GGATACATCAGGATATAATGACAGAGTGAAGTGAAGAGCGATGTCAGCTTAATCCAAT 331  
| | | | | : : : : :  
QY 63 ePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleValTh 83  
| | | | | : : : : :  
Db 332 ATTTCATTTGACTTTAAATAGTAGTATTTAGGGAGGATAGAGTAAAGTATAGTAAC 391  
| | | | | : : : : :  
QY 83 rGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrpIl 103  
| | | | | : : : : :  
Db 392 CCAGAATGAATATATGATATAATCTATGATGAAAGTTTACGATTAGTTTGGAT 451  
| | | | | : : : : :  
QY 103 eArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerValLysAs 123  
| | | | | : : : : :  
Db 452 TAGAATAAATAAGTGAAGTAAATTTACCTGGATATATATAATGATAGTGTAAAAA 511  
| | | | | : : : : :  
QY 123 nAsnSerGlyTrpSerIleGlyIleSerAsnPheLeuValPheThrLeuLysGlnAs 143  
| | | | | : : : : :  
Db 512 TAACCTCAGGTCGAGTATAGGTATATAGTAATTTTGTAGTATTTACTTTAAACAAA 571  
| | | | | : : : : :  
QY 143 nGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTy 163  
| | | | | : : : : :  
Db 572 TGAAGTAGTGAACAAAGTATAATTTAGTATATATATCAATATGCTCTCGATA 631  
| | | | | : : : : :  
QY 163 rAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIl 183  
| | | | | : : : : :  
Db 632 CATAAATGGTTTTTGTAACTGTTACTAACCAATATGATGGGAATATGAGATTTAT 691  
| | | | | : : : : :  
QY 183 eAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLy 203  
| | | | | : : : : :  
Db 692 AAATGGAATTAATAGATATCTATAAAGTTAAAGAACTAACTGGAATTAATTTTACAA 751  
| | | | | : : : : :  
QY 203 sThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs 223  
| | | | | : : : : :  
Db 752 AACTATAACATTTGAATTAATAAATCCAGATACCGGTTTGTACTTACTTCAGATTCTGA 811  
| | | | | : : : : :  
QY 223 pAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAs 243  
| | | | | : : : : :  
Db 812 TAACATCAATATGTTGATGAAGATTTTATATATTTGCTAAAGAAATTAGATGTAAGA 871  
| | | | | : : : : :  
QY 243 pIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAs 263  
| | | | | : : : : :  
Db 872 TATATATATATTTATAGTTTCAATATATCTAATGTTGTAAAGATTTATTTGGGAAA 931  
| | | | | : : : : :  
QY 263 nAspLeuArgTyrAsnLysGluTyrMetValAsnIleAspTyrLeuAsnArgTyrMe 283  
| | | | | : : : : :  
Db 932 TGATTTAAGATATAATAAGAAATATATATGTTAATATAGATTTATTAATAGATATAT 991  
| | | | | : : : : :  
QY 283 tTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAspPheAsnGl 303  
| | | | | : : : : :  
Db 992 GTAGCGAACTCAGCAAAATTTTAAATACAGTGAATATAATGACTTCAATGA 1051  
| | | | | : : : : :  
QY 303 uGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGl 323  
| | | | | : : : : :  
Db 1052 AGGATATAAATTAATAAAGAAATCAGAGAAATACAAATGATAGTACGAGG 1111  
| | | | | : : : : :  
QY 323 yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLy 343  
| | | | | : : : : :  
Db 1112 AGGAGATATTTTATATTTTGTATGACAAATTAATAACAAACATATAATTTGTTATGAA 1171  
| | | | | : : : : :  
QY 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363  
| | | | | : : : : :  
Db 1172 GATGAACTATGTCAGATATATCATAGTACTGACATATATATGCTATAGGTTTAA 1231  
| | | | | : : : : :  
QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383  
| | | | | : : : : :  
Db 1232 AGAACAACAAGGATATAAATGATATATATATATATATATATATATATATATATAT 1291  
| | | | | : : : : :  
QY 383 rTyrTyrTyrAlaSerGlnIlePhePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIl 403  
| | | | | : : : : :  
Db 1292 TTATATATACGATCTCAATATTTAATCAAAATTTAATGAGAAAATATTTCTGCAAT 1351  
| | | | | : : : : :  
QY 403 eCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTrpArgHisAsnTyrLe 423  
| | | | | : : : : :  
Db 1352 ATGTTCAATAGTAGTACTTATCGTTTAGACTTTGGAGGTGATTGGTATAGACAAATTTAT 1411  
| | | | | : : : : :  
QY 423 uValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuLysSerThrThrHistr 443  
| | | | | : : : : :  
Db 1412 GTGGCTACTGTGAGCAGAGAAATTTAGCTTCATTATTAGAACATCACTCACTCATG 1471  
| | | | | : : : : :  
QY 443 pGlyPheValProValSerGlu 450  
| | | | | : : : : :  
Db 1472 GGGTTTGTACCTGTAAGTGA 1493  
| | | | | : : : : :  
RESULT 13  
US-08-704-159-61  
; Sequence 61, Application US/08704159  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; TITLE OF INVENTION: Botulinum Neurotoxin  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/704,159  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 61:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1502 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 108..1493  
US-08-704-159-61  
Alignment Scores:  
Pred. No.: 8,08e-223 Length: 1502  
Score: 2319.00 Matches: 439  
Percent Similarity: 98.66% Conservative: 3  
Best Local Similarity: 97.99% Mismatches: 5  
Query Match: 96.99% Indels: 1  
DB: 11 Gaps: 0  
US-09-910-186a-10 (1-450) x US-08-704-159-61 (1-1502)

QY 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyr 23  
Db 152 CATATCGAAGTCTCATATAGTGGTACATGGCTTTTATTAAAGATATATTAATGAATA 211  
QY 23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43  
Db 212 TTTCAATATATTAATGATTCATAAATTTGAGCTTACAAAACAGAAAAAATACTTACT 271  
QY 43 rAspThrSerGlyTyrAsnAlaGluValSerGluGluGluValAspValGlnLeuAsnProI 53  
Db 272 GGATACATCAGATATATATGAGAAGTGGTGAAGAAGCGATGTTTCAGCTTAATCCAT 331  
QY 63 ePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleValTh 83  
Db 332 ATTTCATTTGACTTTAAATTAGTAGTTCAGGGGAGGATAGAGGTAAGTTATAGTAAC 391  
QY 83 rGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheThrI 103  
Db 392 CCAGATGAAATATATGTATATATCTATGTATGAAGTTTAGCATTAGTTTGGAT 451  
QY 103 eArgIleAsnLysTyrPheValSerAsnLeuProGlyTyrThrIleIleAspSerValLysAs 123  
Db 452 TAGATAATAAATGGTGAAGTAAATTTACCTGGATATCTATATATGATAGTCTTAANA 511  
QY 123 rAsnSerGlyTyrPheIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAs 143  
Db 512 TAACTCAGGTGGATATAGTATATATAGTAAATTTTATAGTATTTTACTTTAAACAAA 571  
QY 143 nGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaProGlyTyr 163  
Db 572 TCAAGTAGTAGAACAAATATATATTTAGTATGATATATCAATAAATGCTCTGGATA 631  
QY 163 rAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrI 183  
Db 632 CAATAAATGGTTTTTGTAACTGTACTAACAATATGATGGAAATATGAAGATTTATAT 691  
QY 183 eAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerL 203  
Db 692 AAATGGAATATATAGATATATATAAGTTAAAGAACTAACTGGAATTAATTTAGCAA 751  
QY 203 sThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs 223  
Db 752 AACTATACAITTTGAATAAATAAATAATCCAGATACCGGTTGATCTACTCAGATTCGA 811  
QY 223 rAsnIleAsnMetTyrIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAs 243  
Db 812 TAACATCAATATGTGATAAGAGATTTTATATATTTGCTAAAGAAATAGATGGTAAGA 871  
QY 243 pIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAs 263  
Db 872 TATTAATATATTTTAACTGTTGCAATATACUAATGTTGTAAGAAATATTTGGGAAA 931  
QY 263 nAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrIleAsnArgTyrMe 293  
Db 932 TGATTTAAGATATAATAAAGAAATATTTATGTTTAAATAGATATTTTAAATAGATAT 991  
QY 283 tTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnG 303  
Db 992 GTATGGAACTACGACAAATGTTTTTATACACGTAGAAATAATGACTTCATGA 1051  
QY 303 uGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 323  
Db 1052 AGGATATAAATATATAAATAAAGAAATCAGAGAAATACAAATGATAGTACTAGATCAGG 1111  
QY 323 yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLy 343  
Db 1112 AGGAGATATTTATTTTGTATGACAAATTAATAACAGCATATATTTTGTATGA 1171  
QY 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363  
Db 1172 GAATGAAACTATGTATGAGATATATCATAGTACTGAAGATATATATCTAGTGTAAAG 1231

QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383  
Db 1232 AGAACAAACAAAGATATAATGATATATATATATTTCAATPACCAACCAATGAATATAC 1291  
QY 383 rTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyT 403  
Db 1292 TTAUTATTAGCATCTCAATATTTAAATCAAATTTAATGAGAAAAATATTTCTGGAAT 1351  
QY 403 eCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyrLe 423  
Db 1352 ATGTTCAATAGTACTTATCGTTTGTAGACTTGGAGTGATTTGGTATAGACACATATTT 1411  
QY 423 uValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHistr 443  
Db 1412 GGTGCCACTGTGAAGCAAGAAATATGCTTCATTTATAGTAATCAACATCAACTCATTC 1471  
QY 443 pGlyPheValProValSerGlu 450  
Db 1472 GGGTTTGTACCTGTAAGTAA 1493  
RESULT 14  
US-10-271-012-61  
; Sequence 61, Application US/10271012  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; Botulinum Neurotoxin  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/271,012  
; FILING DATE: 15-Oct-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/704,159  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 61:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1502 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 108..1493  
; SEQUENCE DESCRIPTION: SEQ ID NO: 61:  
US-10-271-012-61  
Alignment Scores: 8.09e-223 Length: 1502  
Pred. No.: 2319.00 Matches: 439  
Score: 98.66% Conservative: 3  
Percent Similarity: 97.99% Mismatches: 5  
Best Local Similarity:

Query Match:	96.99%	Indels:	1
DB:	43	Gaps:	0
US-09-910-186a-10 (1-450) x US-10-271-012-61 (1-1502)			
QY	4	ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleLeuAsnGluTy	23
DB	152	CCATATCGAAGTCGTCATATGCGTAGCATGCTTTATTAAGATATAATTAATGAATA	211
QY	23	rPheAsnAsnIleAsnAspSerIleLysLeuSerLeuGlnAsnArgLysAsnThrLeuVa	43
DB	212	TTTCATATATATATGATTCACAAATTTTGAGCCACAAACAGAAAATACTTAGT	271
QY	43	laspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIl	63
DB	272	GGATACATCAGGATATAATGCAAGAGTCAGTGAAGGCGATGTCAGCTTAATCAAT	331
QY	63	ePheProPheAspPheLysLeuGlySerSerClyClyAspArgGlyLysValIleValTh	83
DB	332	ATTTCCATTTGACTTTAAATTAGGTAGTTCAGGGAGGATAGAGTAAAGTATAGTAAC	391
QY	83	rGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTyrIl	103
DB	392	CCAGAATGAATAATTTGATATAATTCATGATGAAAGTTTAGCATAGTTTTCGAT	451
QY	103	eArgIleAsnLysTyrValSerAsnLeuProGlyTyrThrIleleaspServallYsAs	123
DB	452	TAGAATAAATAATGGTAAGTAATTTACCGGATATATCTAATGATAGTGTAATAAA	511
QY	123	nAsnSerGlyTyrPsrIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAs	143
DB	512	TAACCTCAGCTGGAGTATAGGTATTATTAGTAATTTTATGATTTTAAACAAAA	571
QY	143	nGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTy	163
DB	572	TGAAGTATGTAACAAAGTATAATTTAGTATGATATATCAATAATGCTCCTGGATA	631
QY	163	rAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIl	183
DB	632	CAATAATGCTTTTGTAACTGTTACTAACATATGATGGGAATATGAGATTTATAT	691
QY	183	eAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyLeuAsnPheSerLy	203
DB	692	AAATGAAATTAATAGATACATATAAAGTTTAAAGAACTAACTGGAATTAATTTAGCA	751
QY	203	sThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs	223
DB	752	AACTATAACATTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	811
QY	223	pAsnIleAsnMetTyrIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAs	243
DB	812	TAACATCAATATGTGGATAAGAGATTTTATATATTTTCTAAAGAAATAGATGTAAGA	871
QY	243	pIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTTPGlyAs	263
DB	872	TATTAATATATTTTATAGCTTGCATATATCTAATGTTGTAAGATTAATTTGGGAAA	931
QY	263	nAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMe	283
DB	932	TGATTTAAGATATAATAAAGATATAATGTTAATATAGATTTATTATATAGATATAT	991
QY	283	tTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGl	303
DB	992	GTATGGGAATCAGCAAAATTTGTTTAAATACAGCTAGAAATTAATATGACTTCAATGA	1051
QY	303	uGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGl	323
DB	1052	AGGATATAAATTAATAAAGAAATCAGAGGAATACAAATGATAGTACAGGACGAGG	1111
QY	323	yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLy	343
DB	1112	AGGAGATATTTTATATTTTATGATGACAAATTAATAACAAAGCATATATTTGTTTATGA	1171

QY	343	sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr	363
DB	1172	GAATGAACACTATGTATGAGATATATCATAGTACTGAGATATATATGCTATAGTTTAA	1231
QY	363	gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPrometAsnAsnTh	383
DB	1232	AGAACAACAAAGGATATAAATATATATATATATATATATATATATATATATATATAT	1291
QY	383	rTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIl	403
DB	1292	TTATATTACGATCTCAAT	1351
QY	403	eCysSerIleGlyThrTyrArgPheArgLeuGlyAspTyrTyrArgHisAsnTyrLe	423
DB	1352	ATGTTCAATAGTACTTATCGTTTAGACTTGGAGTGGAGTGGATGATGATGATGATG	1411
QY	423	uValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHisTr	443
DB	1412	GGTGCTACTGGAAGCAAGAAATTTATGCTTCATTTATAGATCAACATCACTCATTTG	1471
QY	443	pGlyPheValProValSerGlu 450	
DB	1472	GGTTTGTACCTGTAAGTGAA 1493	
RESULT 15			
US-09-611-419A-11			
; Sequence 11, Application US/09611419A			
; GENERAL INFORMATION:			
; APPLICANT: Smith, Leonard A.			
; APPLICANT: Byrne, Michael P.			
; APPLICANT: Middlebrook, John L.			
; APPLICANT: Lapenotiere, Hugh			
; APPLICANT: Clayton, Michael A.			
; APPLICANT: Brown, Douglas R.			
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM			
; FILE REFERENCE: A33626 067252.0105			
; CURRENT APPLICATION NUMBER: US/09/611.419A			
; CURRENT FILING DATE: 2000-07-05			
; PRIOR APPLICATION NUMBER: PCT/US00/12890			
; PRIOR FILING DATE: 2000-05-12			
; PRIOR APPLICATION NUMBER: 60/133,865			
; PRIOR FILING DATE: 1999-05-12			
; PRIOR APPLICATION NUMBER: 60/133,866			
; PRIOR FILING DATE: 1999-05-12			
; PRIOR APPLICATION NUMBER: 60/133,867			
; PRIOR FILING DATE: 1999-05-12			
; PRIOR APPLICATION NUMBER: 60/133,868			
; PRIOR FILING DATE: 1999-05-12			
; PRIOR APPLICATION NUMBER: 60/133,869			
; PRIOR FILING DATE: 1999-05-12			
; PRIOR APPLICATION NUMBER: 60/146,192			
; PRIOR FILING DATE: 1999-07-29			
; NUMBER OF SEQ ID NOS: 42			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 11			
; LENGTH: 1374			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Synthetic construct based on BONTA HC			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (10)...(1362)			
US-09-611-419A-11			
Alignment Scores:			
Pred. No.:	7, 94e-77	Length:	1374
Score:	868.50	Matches:	197
Percent Similarity:	62.06%	Conservative:	86
Best Local Similarity:	43.20%	Mismatches:	145
Query Match:	36.32%	Indels:	28
DB:	23	Gaps:	11

```

1036 ATCATCCGTTGACACCGACACCATCTACGCCACCGAGGTTGTTGCTTCCCAAGACTGT 1095
QY 357 IleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln 376
Db 1096 GTCACGCGCTGAACCTGACCTCCCAACCTGGTAACTACGATATCGGT--ATCTTCTCC 1152
QY 377 IleGlnProMetAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsn 396
Db 1153 ATCAAGAATCTGCTCCCAAGAACAAGTACTGCTCCAGATCTTC--TCTCTCTCCGT 1209
QY 397 GlyGluAsnIleSerGlyIleCysSerIle--GlyThrTyrArgPheArgLeuGlyGly 415
Db 1210 ---GAGACACCATCTGCTGGCCGACATCTACAGCCTTGGCGTTTCTCC----- 1257
QY 416 AspTyrTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu 435
Db 1258 -----TTCAAGAAGCGCTACACTCTCTGCGCGTACC-----AACTAGAGACCAAG 1305
QY 436 LeuGluSerThrSerThrHisTyrGlyPheVal-ProValSerGlu 450
Db 1306 CAGCTGTCCACCTCTCTCTTCTGGAGTTTCACTCTCCCGTGACCCAG 1351

```

Search completed: November 7, 2002, 20:04:45  
Job time : 3179 secs

```

US-09-910-186A-10 (1-450) x US-09-611-419A-11 (1-1374)
QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsn 21
Db 49 ACCATCCCATTCACATCTTCTTACACCACTCTCTTGTGAGGACATCATCAAC 108
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 109 GAGTACTTCACTCCATCAACGACTCCAGATCTTGTCTCTCCAGAACAGAGAGAGCC 168
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluArgValGlnLeuAsn 61
Db 169 TTGTCGACACTCCGGTTTACACCCGAGGTACAGTGGGTGACAGCTCCAGTTGAC 228
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValile 81
Db 229 ACCATCTACACCAAGACTTCAAGTTGCTCTTCCGCTGAC-----AAGATCATC 279
QY 82 ValThrGluAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 280 GTCAACTTGAAACAACAACATCTTGTACTCCGCCATCTACGAGAACTCTCTGTCTCTTC 339
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyrThrIleIleAspSer 120
Db 340 TGGATCAAGATCTCCAAAGGACTTGACCAACTCCCAACAGGAGTACACCATCATCAACTCC 399
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
Db 400 ATCGACGACAGACTCCGGTTGGAGTTGTATCCGATACCGTATACATCGAGTGGATCTTG 459
QY 141 LysGluAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
Db 460 CAGGACGTCAACCGTAAGTACAGTCTTGATCTTCGACTACTCCGAGTCTCTCTCCAC 519
QY 161 ProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet 179
Db 520 ACCGGTTACACCAACAAAGTGTCTCTGTCACCATCCCAACAACATCATCGGTATCATG 579
QY 180 LysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199
Db 580 AAGTTGTACATCAACGGTGAGTGAAGCAGTCCCAAGAGATCGAGGAGCTTGACGAGGTC 639
QY 200 AsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr 219
Db 640 AAGCTGGACAACACCATCTCTCTGTTGTCGACGAGAGACATCGAC----- 684
QY 220 SerAspSerAsnAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeu 239
Db 685 -----GAGAACACAGATGTGTGGATCCGCTGACTTCAACATCTCTCCAGAGAGCTG 735
QY 240 AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAsp 259
Db 736 TCCAAGGAGGACATCAACATCTCTACGAGGTCAGATCTCTGAGGAGCGTCTCATCAAGGAC 795
QY 260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279
Db 796 TACTGGGGTACCCACTGAAGTTCGACCCGAGTACTACATCATCAACGACAGCACTATC 855
QY 280 AsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn 299
Db 856 GACCGTTACATGCCGCCAGAGTCCCAACGCTCTGTCTGCTGCTGCTGCTGCTGCTGCT 915
QY 300 AspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThr 319
Db 916 AAGCTGTACACCGGTAAACCTTATACCATCAAGTCCGCTCTCCGACAGAACCCCTTACTCC 975
QY 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339
Db 976 CGTATCTTGAACGGTGACAACTATCTCTGTCACATGCTGTGATCAACTCCGCTGAGTACATG 1035
QY 340 LeuPheMetLysAsnGluThrMetTyrAla-----AspAsnHisSerThrGluAsp 356

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 17:52:07 : Search time 31 Seconds  
(without alignments)  
1596.894 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTFPFNFSTNNLLKXDI.....NYASLLESTTHGWFVPVSE 450

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 193892 seqs, 55004114 residues

Total number of hits satisfying chosen parameters: 387784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+pn2n.model -DEV=xlh

DB=US-09-910-186A-10

DB-PENDING=Patents.NA.New -QFMT=fastap -SUFFIX=rpnp -MINMATCH=0.1 -LOOPCL=0

LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.ccd1

LIST=45 -DOALIGN=200 -THR SCORE=100 -THR MAX=100 -THR MIN=0 -ALIGN=15

MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

USER=US09910186.ecgn1\_l3 -runat\_04112002\_111619\_20924 -NCPU=6 -ICPU=3

NO.XLPY -NO.MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120

WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7

YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Pending Patents.NA.New.\*

2: /cgn2\_6/ptodata/1/pna/PCT\_NEW\_COMB.seq.\*

3: /cgn2\_6/ptodata/1/pna/US06\_NEW\_COMB.seq.\*

4: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq.\*

5: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq.\*

6: /cgn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq.\*

7: /cgn2\_6/ptodata/1/pna/US10\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122	5.1	2739	US-10-092-411A-156	Sequence 156, App
2	121.5	5.1	4008	US-10-092-411A-879	Sequence 879, App
3	119.5	5.0	2793	US-10-092-411A-2477	Sequence 2477, App
4	117	4.9	2076	US-10-092-411A-1838	Sequence 1838, App
5	116.5	4.9	2376	US-10-240-435-1411	Sequence 1411, App
6	116.5	4.9	3456	US-10-092-411A-405	Sequence 405, App
7	114.5	4.8	1494	US-10-092-411A-2277	Sequence 2277, App
8	112	4.7	6070	US-10-240-485-131	Sequence 131, App
9	111	4.6	6713	US-10-240-485-79	Sequence 79, App
10	108.5	4.5	5979	US-10-240-453-25	Sequence 25, App
11	108	4.5	2931	PCT-US02-05068-97	Sequence 97, App

c	12	108	4.5	3396	5	US-09-721-456-640	Sequence 640, App
c	13	106	4.4	3033	6	US-10-092-411A-2341	Sequence 2341, App
c	14	105.5	4.4	6725	6	US-10-240-485-107	Sequence 107, App
c	15	105	4.4	9265	6	US-10-240-454-49	Sequence 49, App
c	16	105	4.4	30549	6	US-10-092-411A-322	Sequence 322, App
c	17	104.5	4.4	2088	6	US-10-092-411A-1504	Sequence 1504, App
c	18	104.5	4.4	3396	5	US-09-721-456-641	Sequence 641, App
c	19	102.5	4.3	3279	5	US-09-721-456-109	Sequence 109, App
c	20	102	4.3	11091	6	US-10-092-411A-2243	Sequence 2243, App
c	21	101	4.2	12405	6	US-10-240-453-44	Sequence 44, App
c	22	101	4.2	1584	6	US-10-092-411A-1628	Sequence 1628, App
c	23	100	4.2	2415	6	US-10-092-411A-2381	Sequence 2381, App
c	24	99.5	4.2	2013	6	US-10-092-411A-103	Sequence 103, App
c	25	99.5	4.2	2028	6	US-10-092-411A-1710	Sequence 1710, App
c	26	99.5	4.2	4199	6	US-10-145-087A-497	Sequence 497, App
c	27	99.5	4.2	4199	6	US-10-143-031A-497	Sequence 497, App
c	28	99.5	4.2	4199	6	US-10-145-092A-497	Sequence 497, App
c	29	99.5	4.2	4199	6	US-10-182-322A-497	Sequence 497, App
c	30	99.5	4.2	4199	6	US-10-165-038A-497	Sequence 497, App
c	31	99.5	4.2	4199	6	US-10-165-353-497	Sequence 497, App
c	32	99.5	4.2	4199	6	US-10-170-481A-497	Sequence 497, App
c	33	99.5	4.2	4199	6	US-10-172-039A-497	Sequence 497, App
c	34	99.5	4.2	4199	6	US-10-145-016A-497	Sequence 497, App
c	35	99.5	4.2	4199	6	US-10-145-088A-497	Sequence 497, App
c	36	99.5	4.2	4199	6	US-10-145-329A-497	Sequence 497, App
c	37	99.5	4.2	4199	6	US-10-165-353A-497	Sequence 497, App
c	38	99.5	4.2	5464	6	US-10-240-454-38	Sequence 38, App
c	39	99.5	4.2	13700	6	US-10-085-198-21	Sequence 21, App
c	40	99	4.1	2871	6	US-10-092-411A-1615	Sequence 1615, App
c	41	98.5	4.1	1269	6	US-10-092-411A-2393	Sequence 2393, App
c	42	98.5	4.1	2421	6	US-10-092-411A-1477	Sequence 1477, App
c	43	98.5	4.1	2442	6	US-10-268-229-10	Sequence 10, App
c	44	97.5	4.1	1287	6	US-10-092-411A-105	Sequence 105, App
c	45	97.5	4.1	1541	6	US-10-240-485-193	Sequence 193, App

#### ALIGNMENTS

RESULT 1  
US-10-092-411A-156  
; Sequence 156, Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092-411A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 156  
; LENGTH: 2739  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-156

Alignment Scores:  
Pred. No.: 0.000269 Length: 2739  
Score: 122.00 Matches: 96  
Percent Similarity: 35.05% Conservative: 72  
Best Local Similarity: 20.60% Mismatches: 154  
Query Match: 5.10% Indels: 144  
DB: 6 Gaps: 23

US-09-910-186A-10 (1-450) x US-10-092-411A-156 (1-2739)

QY 3 IleProPheAsnIlePheSerTyrThrAsnAsn----- 13  
:|||||: :  
:|||||: :  
:|||||: :

1120 GNTCCATTGTGATCAAAATGCCATTACAGCCTCTTTAAAGAGCCATAATCCTAATATCTAT 1179

14 SerLeuLeuLysAspIleleAeSnglufyrPheAsnAsnIleAsnSerLysIleLeu 33

1180 ACTTTGTTTACAAATGATA-----GAATACAAAGGG-----AGAGAAATATGAATTATTA 1227

34 SerLeuGlnAsnArclLysAsnThrLeuValAspThrSerGlyItyrAsnAla-----Glu 51

1228 GTCAGAGAAGATTCAAAAGAACTTTGAAGAAATGAAGAAATTTATACACCAAGAAAGAA 1287

52 ValSerGluGluGlyAspValGlnLeuAsnProIlePheProPheAspPheLysLeuGly 71

1288 ATTGACCAAGTATGGTGTGTA----- 1308

72 SerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyr--- 90

1309 -----AACGAATATCAATGAATAATATAATATATTTGTTATTATAAG 1350

91 -----AsnSerMetYrGluSerPheSerIleSerPheTrpIleArgIleAsnLys 107

1351 CATCGACCAATAGGGAATACATTCATTTGGTGATAAATATTTATATGAAATAGAG 1410

108 TrpValSerAsnLeuProGlyTyrThrIleIleAspSerValLysAsnAsnSerGlyTrp 127

1411 CAGAGTATT-----ATGAGTATATAAAGTCATTAACATTTGTCAAATGAA 1458

128 SerIleGlyIleIleSerAsnPheLeuValPheThrLeu----- 140

1459 AGCGTAGGGGTATTCAAATTCAGTAGAGTTTGGCTAAATATGAATATACATCGTTC 1518

141 -----LysGlnAsnGluAsp-----SerGluGlnSerIleAsnPhe 152

1519 GATTGTAAAGAAACAGGAATACTCTTAAGATGTTTCCAATATCAAAATATCATGATC 1578

153 SerTyrAspIleSerAsnAsnAlaProGlyTyrAsnLysTrpPhePheValThrValThr 172

1579 GATAC-----GGAGCTCTGGACAGAGAAATCAACTTAATTAACCATATT 1626

173 AsnAsnMetMetGlyAsnMetLysIleTyrIle-----AsnGlyLysLeuIleAsp 189

1627 TGCATTTTTCTATGATAAGATGTTATCGTCATAGCAAAATCTAATCTACGAGTAGAC 1686

190 ThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle 209

1687 AATATTAAAGAAATA-----ATCAAAATATCTACATTAAGAAGCTCTACTATT 1734

210 AsnLysIleProAspThrGlyLeuIleThrSerAspSerAsnIleAsnMetTrpIle 229

1735 TCTAAA-----TTTTTATATAATGATAAGAAAGATGACTTGCTTAATA 1779

230 ArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn 249

1780 ATCGAT-----GAGCGAGGTACAGTTAGTAATAGGACATGAATGAATCTTTGAA 1830

250 SerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLys 269

1831 AACACGAATTTGTAATTATA-----TTAATGTGCGGTGAT 1866

270 GluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsn----- 286

1867 AATTATCAAAATCGAATCGATAGATTTTGGAAATTTGGTTTCGAATGCGCAAGATGTTTC 1926

287 SerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGluGlyTyrLys 306

1927 TCAAAAATATATATC-----AACGAACTAATCATGATGATAT--- 1962

307 IleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyLysAspIle 326

1963 -----CGACTAAATATGATGATTA 1983

327 LeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThr 346

1984 CTTTACTTTTGGAAATCTGTTATAGAAAAAAGAGTAAATTA-----AATGAAAT 2034

```

QY 347 MetTyrAlaAspAsnHisSerThr----- 354
   :: :: :: :: :: :: :: :: :: :: ::
Db 2035 ATTAATATGAATAAATAATTTCTACAGATAGATGAAAGTATATTATTAATGAATCAATAAA 2094
QY 355 -----gluaspiletyralailleglyleuargGlu--- 364
   :: :: :: :: :: :: :: :: :: :: ::
Db 2095 GATGAATATTCTTTGTTTAAATTCAGATGGTATATATGGTATCATATANTAGA 2154
QY 365 -----GlnThrLysaspIleAsnAspAsnIleIlePheGlnIleGlnPro----- 379
   ||| :: :: :: :: :: :: :: :: :: :: ::
Db 2155 TTATTACAAGCAATAATAAATAATGATTCTCTAATTTGGGTGTGAAAGAAATAATAAGTT 2214
QY 380 -----MetAsnAsnThrTyrTyrTyrTyrAlaSerGlnIlePheLysSerAsn 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2215 GGTGATCTATTCTATTTAATGAACAACAACTAATAGTACTCACCATACTTTT---AATAAT 2271
QY 395 PheAsnGlyGluAsnIle 400
   ||| ||| |||
Db 2272 TTAAAAGGGTCAATAATT 2289

RESULT 2
US-10-092-411A-879
; Sequence 879, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STATA
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 879
; LENGTH: 4008
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-879

Alignment Scores:
Pred. No.: 0.000521 Length: 4008
Score: 121.50 Matches: 97
Percent Similarity: 34.83% Conservat: 74
Best Local Similarity: 19.76% Mismatches: 193
Query Match: 5.08% Indels: 127
DB: 6 Gaps: 23

US-09-910-186A-10 (1-450) x US-10-092-411A-879 (1-4008)
QY 10 TyrThrAsnAsnSerLeuLeuLysaspIleIleAsnGlnLutyrPheAsnAsnIleAsnAsp 29
   ||| ||| :: :: ||| ||| ||| ||| ||| ||| ||| |||
Db 2698 TAGCTTAACCGTACTTCCTT-----ATCAATAACAACGCTACTCAAGTAATAAC 2748
QY 30 SerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsn 49
   :: :: :: :: ||| ||| ||| ||| ||| ||| ||| |||
Db 2749 ACCTATGTACTACTTCAAGTACGACACAGTAACTACTCAATTA----- 2790
QY 50 AlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePheProPheAspPheLys 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2791 -----GGATGGGTAAACATTATATGTGTGACAACTCAAAAT----- 2826
QY 70 LeuGlySerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleVal 89
   :: :: :: :: ||| ||| ||| ||| ||| ||| ||| |||
Db 2827 ATCGGAACAACAACACTGCTCTAGTAAATATTCAAGTAAACCTACAAATAATGGTCTA 2886
QY 90 TyrAsnSerMetTyr----- 94
   ||| :: ::
Db 2887 TATTCATTGCTTTGGGGTACTAAACAACCAACAATTAAGTACACCTTAATACGCTAGCTAAT 2946

```

Copied from 10910186 on 05-05-2004



```

Qy 150 ileAsnPheSerTyraSpilleSerAsnAlaProGlyTyraSlnYsrPhePheVal 169
Db 1069 GCTACAGTCTTATGACACACAAATAAACAATACCTACACCTTTACAGATTAATGA 1128
Qy 170 ThrValThrAsnAsnMetGlyAsnMetLysile-----Tyrile----- 183
Db 1129 GATAAATATGAAATATATAAGCGCCTTAAATTAATACATCATACATATGATAAATCAAG 1188
Qy 184 -----AsnGlyLysLeuileAspThrileLysValLysleuThrGlyleAsn 200
Db 1189 GTTCAATATACACTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1248
Qy 201 PheSerLysThrilePheGluileAsnLysileProasp-----ThrGlyLeuile 218
Db 1249 -----AAAACAATTCGGTGTGATATCAAAACCTAACGAAATCGGACTGCTAACCTT 1302
Qy 219 ThrSerAspSerAsnAsnleAsn-----MetThrile 229
Db 1303 CAAGATGTTTCACAAATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1362
Qy 230 ArgAspPheThrilePheAlaLysGluLeuAspGlyLysAspileAsnleLeuPheAsn 249
Db 1363 AACCTCTTCCTTATTCAGCCAAAGAA-----ACAAATGTAATATATTCAGGGAAT 1413
Qy 250 SerLeuGlnThrAsnValValLysAspTyrrPheAsnAspLeuArgTyraSlnLys 269
Db 1414 GCGGATGAGGTTCACAAATATTCAGCAT-----AGTACAATCAATTAAGTTTATAAG 1467
Qy 270 GluTyrrMetValAsnleAspTyrrLeuAsnArg-----TyrMetTyrrAlaAsnSerArg 288
Db 1468 GTTGAGATAATCAAAATTCACGATAGTACAGAAATTTATGATGATGATGATGATGAT 1527
Qy 289 GlnleValPheAsnThrArgAsnAsnAsnAspPheAsnGluGlyTyrrLysilele 308
Db 1528 GATGTC-----ACAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1554
Qy 309 ileLysArgileArgGlyAsnThrAsnAspThrArgValArgGlyLysAspile----- 326
Db 1555 -----TAGGAATAATTAAGTGTGATGATGATGATGATGATGATGATGATGATGAT 1602
Qy 327 LeuTyrrPheAspMetThrile-----AsnAsnLysAlaTyrrAsnLeuPhe 341
Db 1603 CATATATATTAAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1662
Qy 342 MetLysAsnGluThrMetTyrrAlaAsp-----AsnHisSerThrGluAspileTyrrAlaile 360
Db 1663 CAGCAAACTGTGCACATCCAAACGACTATTAATGATGATGATGATGATGATGATGAT 1713
Qy 361 GlyLeuA-gGluGlnThrLysAspileAsnAspAsnleilePheGlnleGln----- 378
Db 1714 -----AGAACACGATCCTATGATGATGATGATGATGATGATGATGATGATGAT 1761
Qy 379 -----ProMetAsnAsnThrTyrrTyrrTyrrTyrrTyrrTyrrTyrrTyrrTyrr 392
Db 1762 CAAGCAGAGGTGACTTCCTCCCTGAAACAACTTATAAATCGGAGATTACGTATGGGA 1821
Qy 393 -----SerAsnPheAsnGlyGluAsnleleSerGlyile 403
Db 1822 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1881
Qy 404 CysSerileGlyThrTyrr 409
Db 1882 TTGGTAACCTTCACGAT 1899

```

RESULT 4  
 US-10-092-411A-1838  
 ; Sequence 1838, Application US/10092411A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 032796-101  
 ; CURRENT APPLICATION NUMBER: US/10/092,411A

```

; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 1838
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-10-092-411A-1838

Alignment Scores:      Length:      2076
Pred. No.:            0.000661
Score:                117.00
Percent Similarity:   33.50%
Best Local Similarity: 19.43%
Query Match:          4.89%
DB:                   6
Gaps:                 33

US-09-910-186A-10 (1-450) x US-10-092-411A-1838 (1-2076)

Qy 1 MethThrilePheAsnlePheSerTyrrAsnAsnSer----- 14
Db 13 ATGCAATACCAATTAATTA-----CCTACAAACAGTACTGATTAATGAATTAATGT 66
Qy 15 -----LeuLeuLysAspileleAsn 21
Db 67 ACATTACAAAGTAGAACCATCAATATAAAAGGAGAGGTTTAACTACTGAAATATATGAT 126
Qy 22 GluTyrrPheAsnAsnleAsnAspSerLysileleu----- 33
Db 127 GATACATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 186
Qy 34 -----SerLeuGlnAsn-----ArgLysAsnThrLeu-----ValAspThr 45
Db 187 CAAGATAGTATCAAGAACTATAGAGACAAAGAAAGAAATGTTTCTTTAGGATTAAGT 246
Qy 46 SerGlyTyrrAsnAlaGlu-----ValSerGluGlu----- 55
Db 247 AAAAATTGAAATTTAGAAATTAATACCTTTTGTAAACATGATGTTAAGAGGATG 306
Qy 56 -----GlyAspValGlnLeuAsnProilePhePropheAsp 67
Db 307 TCTTATCTAATTTATTTAATCTGAGCAGTAAAGCTTAATAAGATTCGTAAGTTC 363
Qy 68 PheLysLeuGlySerSerGlyGluAspArgGlyLysValleValThrGlnAsnGluAsn 87
Db 363 ----- 363
Qy 88 IleValTyrrAsnSerMetTyrrGluSerPheSerileSerPheTrpIleArgileAsn 106
Db 364 -----TTTAATGAAGTTATCCAAATTTAAATTTCTTACTTGTGATGATATAAATACA 417
Qy 107 -----LysTrpVal-----SerAsnLeuPro-----Gly 114
Db 418 TTAGAAAGACACTGTTTAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 477
Qy 115 TyrThrIleleAspSerValLysAsnAsnSerGlyTrpSerileGlyIleleSerAsn 134
Db 478 TCACATATAGTTTTTGGTGATGATGATGATGATGATGATGATGATGATGATGAT 537
Qy 135 PheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerileAsnPheSerTyrr 154
Db 538 ATGTACATTAATCTAATTAAGTTTCATGATGATGATGATGATGATGATGATGATGAT 597
Qy 155 AspIleSerAsn-----AsnAlaProGlyTyrrAsnLysTrpPhePheValThrVal 171
Db 598 GATATAGAAATTTAGAAAAATATGATGATGATGATGATGATGATGATGATGATGAT 645
Qy 172 ThrAsnAsnMetMetGlyAsnMetLysileTyrrileAsnGlyLysleuLeuAspThrile 191

```



```

Oy 154 TyrAsp-----lleSerAsnAsnAlaProGlyTyrAsnLysTrpPheValThrVal 171
    |||
Db 610 TACAGAGAGTCTCTCCCAAGAGACTGCTGGC-----642
Oy 172 ThrAsnAsnMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIle 191
    |||
Db 643 -----ATAGAGCATCTGCTCCTCTGTTTAACTTCAAGAGACATACAGTA 690
Oy 192 LysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLys 211
    |||
Db 691 -----GCAGACTGTGCTCTAATTAATAAATATTATGGAACGAAATCCT 735
Oy 212 IleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgasp 231
    |||
Db 736 GTTCAGGCTATTCGTTCAGACAGACAGACAGATCCATACACCTTGGGGAAAGAC 795
Oy 232 -----PheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIle 246
    |||
Db 796 CATGAAAAAGATGTTTGAACATATTGTAACACAGTTTTCATCAGTCTGTATCTGTG 855
Oy 247 LeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArg 266
    |||
Db 856 GTACGCGATCTATGACATTTAATGCGTGTGAGAAATATGGGGTGAAGATCTAAGA 915
Oy 267 TyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsn 286
    |||
Db 916 CAT-----TTAATAGTATCGAGA 933
Oy 287 SerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyrLys 306
    |||
Db 934 AGTACAGAGCA-----CCA 948
Oy 307 IleIleIleLysArgIleArgLysThrAsnAspThrArgValArgGlyGlyAspIle 326
    |||
Db 949 CTAATAACAGACCTGATCTCGAACCCTCTTGACACTGTGTTAAGGTTTGGAGATT 1008
Oy 327 Leu-----TyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
    |||
Db 1009 TTAGTAAAGATTCTCTGTACTGAGAACTCAAGGGTTACAAGTTG 1056

```

## RESULT 6

US-10-092-411A-405

Sequence 405, Application US/10092411A

## GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 032796-101

CURRENT APPLICATION NUMBER: US/10/092,411A

PRIOR FILING DATE: 2002-03-07

PRIOR APPLICATION NUMBER: US 09/134,001

PRIOR FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/035,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5676

SEQ ID NO 405

TYPE: DNA

LENGTH: 3456

ORGANISM: Staphylococcus epidermidis

US-10-092-411A-405

## Alignment Scores:

Pred. No.:	0.00153	Length:	3456
Score:	116.50	Matches:	97
Percent Similarity:	34.44%	Conservative:	90
Best Local Similarity:	17.86%	Mismatches:	175
Query Match:	4.87%	Indels:	181
DB:	6	Gaps:	27

US-09-910-186a-10 (1-450) x US-10-092-411A-405 (1-3456)

```

Oy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22
    |||
Db 214 CTACCAAGTTCTCCAGTACCAACAGAGCAATACAGCTATCTGTTTCAGTACCAAGGAA 273
Oy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
    |||
Db 274 -----GACGAAATCAATGCCAATCTTACGCACACAAATATACAAA 315
Oy 43 Val-----AspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspVal 58
    |||
Db 316 TTAACAAAGTACAAACAGTGTAAACAAATGATCGACAAATGTCGATAGAAATCGATA 375
Oy 59 GlnLeuAsnProIlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArg 77
    |||
Db 376 GAATTTAT-----TTCCATCAACAAATTTAATCTAGGGTTTACATTTCAAGATACA 429
Oy 78 GlyLysVal-----lleValThr 83
    |||
Db 430 CCGGCTGTGACTCTAAGCTTGAACGCATCAGTCAGTACAGACGAATTTATGTATACA 489
Oy 84 GlnAsn-----GluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100
    |||
Db 490 AGTAACTGTGTTTATACAGTAGACTATATCATGTCCAATCAGCGTTGACCTTAA 549
Oy 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120
    |||
Db 550 TTTATGAACGCTAATAATGAA-----GTTGGTATACCAATATTATTTTGTGATTAATCAG 603
Oy 121 Val-----Lys 122
    |||
Db 604 ATTATTAACATATGAAGAAGAAATTACATTTGAACCTTTAAATCAACAGATCGAATAA 563
Oy 123 AsnAsnSerGlyTrpSerIleGlyIle----- 131
    |||
Db 664 TCAATCAAGAGCTGGGATATCAAACTTCAAGATCTATTATTCGTTTCAAAAGTTGATCAT 723
Oy 132 -----lleSerAspPheLeuValPheThrLeuLysGlnAsnGlu 144
    |||
Db 724 CCACAGATGAATGTACAACTTTCAATTTCTAGTATTTATGATCAATCAATCGTGA 783
Oy 145 AspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyrAsn 164
    |||
Db 784 TCAACAGAGACTGTATTAAGACAACTCAATTCATTCACGACGACAA-----834
Oy 165 LysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleThrIleAsn 184
    |||
Db 835 -----TACATATACATTCAAATGAATGCAA-----861
Oy 185 GlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThr 204
    |||
Db 862 ---TCTATTCTGTACACCTTCAAAATTAATGAA-----GAACAATTCGAGGAAGCA 909
Oy 205 ---lleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAsp 223
    |||
Db 910 TATATTCAATTTCAACAAATACAA-----GAAGTCAGCGCAGAGACACAA 954
Oy 224 AsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAsp 243
    |||
Db 955 TTGCTCAAT-----GAC 966
Oy 244 IleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsn 263
    |||
Db 967 TCTAATCAATTTAATTTAATTTAAACAGAACGCT-----AAGATATATTAGATAAT 1020
Oy 264 -----AspLeuArgTyrAsnLysGluTyrTyrMetValAsnIle---276
    |||
Db 1021 GCTTATATCATGACGTACGATATCGCGAATCTTTACGGGAATTTATTAGAACATGGCA 1080
Oy 277 ---AspTyr-----LeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThr 294
    |||
Db 1081 ACTGATTTTAAAGTGAATGGATT-----TTTAATAAA 1113
Oy 295 ArgArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGly 314

```

```

Db 1114 AGGAGAAAAAGAGAGAGACAA-----ATCAACGACCTTAATGAG 1155
Qy 315 AsnThrAsnAspThrArgValArgGlyGly-----Asp 325
Db 1156 CGGACCACTCAATTCGACGAGAAAGTAATCAACAAGTACGACCACTTCGTGAAGAT 1215
Qy 326 IleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGlu 345
Db 1216 ATGTCATTTTAACTAGATCAATAATAAATCAATGCGTGAATGAAATATAATCAAA 1275
Qy 346 -----ThrMetTyrAlaAspAsnHisSerThrGlu 355
Db 1276 GAATATGACGTCGTCGTCACCTATATACAGACTATATCAAACTCAAAAGCATTAGC 1335
Qy 356 AspIleTyrAlaIleGlyLeuArgGluInThr-----LysAspIleAsnAspAsnIleIle 374
Db 1336 AACATACAGCTTTTACATATTCAGATGAAGTTATTAAGCTTTGAATAAAAATAGAA 1395
Qy 375 PheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsn 394
Db 1396 AATGAGTCAACACCA-----CTATTTGAAGAAGCT 1425
Qy 395 PheAsnGlyGluAsnIleSerGlyIleCysSer----- 405
Db 1426 GTCAATCATGATACAAAGTTAATGAATATCATCGATGATCAAAATGAAGATAGGATGAATAT 1485
Qy 406 -----IleGlyThrTyrArgPheArgLeuGlyGlyAspTyrIleArgHis 420
Db 1486 GATAGATACATGGAACCTAACACATTAAGGATTCGCTTACATCCACACATACAAACAT 1545
Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440
Db 1546 TACTATATC-----CATTTAGACGATCTTTAGATAAATTAATTTGGAAGACAGAG 1596
Qy 441 ThrIleIleThr 443
Db 1597 ACTCATTTT 1605

```

## RESULT 7

```

Sequence 2277, Application US/10092411A
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-101
CURRENT APPLICATION NUMBER: US/10/092,411A
PRIOR FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: US 09/134,001
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5676
SEQ ID NO 2277
LENGTH: 1494
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-10-092-411A-2277

```

```

Alignment Scores:
Pred. No.: 0.000794 Length: 1494
Score: 114.50 Matches: 90
Percent Similarity: 35.27% Conservative: 56
Best Local Similarity: 21.74% Mismatches: 158
Query Match: 4.79% Indels: 110
Gaps: 21

```

US-09-910-186A-10 (1-450) x US-10-092-411A-2277 (1-1494)

Qy 10 TyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAsp 29

```

Db 300 TACACGTTTAAACGGGACATGAGATGAATAAATCAATGAATTCAGTCGACATAGCAACGAT 359
Qy 30 SerLysIleLeuSerLeuGlnAsnArgLys-----Asn 40
Db 360 TGTACACACGTTGCCCTTAGAATTAGAAGGGTTTAATTTATATCATTTTAATGATTAAAC 419
Qy 41 ThrLeu-----Val-AspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspVa 58
Db 420 ACAATTAGATAAACAATGATGTTTCAGGTTACAAA-----TT 455
Qy 58 LglnLeuAsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGl 78
Db 456 TAATTAGAATTCGATTCGATCAATTAATGTTACGTCGACGACGGACCTCAAAAGC 515
Qy 78 YLysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSe 98
Db 516 TGTCCCTCAACAGCTTAAATAATCATTTAGCCAGTCTAAAGCTGTAAACAAGTTAGG 575
Qy 98 IleSerPheTyrIleArgIleAsnLysIleValSerAsnLeuProGlyTyrThrIleIle 118
Db 576 ATTCGAA-----CAAAATACTGTGTGGCTTTCGGTCTTACCTATATATCATATT-- 624
Qy 118 eAspSerValLysAsnAsnSerGlyTyrPheIle-----GlyIleIleSerAsnPh 135
Db 625 -----TCTGGCTCAGTGTATTTTCCGCGCAGGATGATACAGGAT 665
Qy 135 eLeuValPheThrLeuLys-----GlnAsnGluAspSerGluGlnSerIleAsnPheSe 153
Db 666 CACGTGCAGACTGTTTAAAGATTTCAAACTGATGATATGTTAAACACAAATAAG--AC 722
Qy 153 rTyrAspIleSerAsnAsnAla-----ProGlyTyrAsnLysTyrPhePheValThrVa 171
Db 723 TTATCCCAATCACCATATGTCCTTGTCCCAACAAAGCTTAAGTGGTTAATGATGCGAG 782
Qy 171 lThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly-----Lys 186
Db 783 ATTGACTCAACCATTTTCTTTAGAAAAAATTCGTAGTGGTGTCTAAATATATCACACA 842
Qy 186 sLeuIleAsp-----ThrIleLysValLysGluLeuThrGlyIleAsnPheSerly 203
Db 843 ATTAATTGACCAAGCACTGATTCGTTTACCTGTATATAATCTTTGGTATGACAGA 902
Qy 203 sThrIleThrPheGluIleAsnLysIlePro-----AspThrGl 216
Db 903 AACTGTCTCTCAGTTCTTACAGCCCTCACCTCAATGCTCAAGAAGCGTTTCGATACT-- 960
Qy 216 yLeuIleThrSerAspSerAspAsnIleAsnMetTyrIleArgAspPheTyrIlePheAl 236
Db 961 ---GTTGGAAAAACCAAGTGAATGCGAAGTGAATAAATAAATAATCCCAACGCATATGG 1016
Qy 236 aLys---GluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGluTyrThrAs 255
Db 1017 ACATGGAGAGCTTATTAATAAGTGAAATGTGATG--AATGGTTATTATATATCCCAA 1073
Qy 255 nValValLysAspTyrTyrPglyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValas 275
Db 1074 ATATTAAAGACACATTTGATATGAT---GGGTATTTCAAACTGGAGATATAGCTGA 1130
Qy 275 nIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrAr 295
Db 1131 AATAGAT----- 1137
Qy 295 gArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAs 315
Db 1138 -----GATGAAGGTTTACGTCATAATATATATGATCGG----- 1167
Qy 315 nThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAs 335
Db 1168 -CGCAAGATTTGATTATAGTGGTGGAGAGATATTTAT----- 1206
Qy 335 nLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGl 355

```







QY 391 eLysSerAsn-----PheAsnGly-----GluAsnIleSerGlyLeCysSerIl 406  
||| ||| ||||| :||||:  
Db 1887 AAAAAAACAACATCTCTTTAAATAATAATAATAATAAAAAACTTA-----AT 1843  
||| ||| ||||| :||||:  
QY 406 eGlyThrTyr-----ArgPheArgLeuGlyGlyAspTpTYrArgHisAsp----- 421  
||| ||| ||||| :||||:  
Db 1842 AACTACATACAAAATAAAATAAAATTAACCTTCACCACCTCTCTCCCAATTAA 1783  
||| ||| ||||| :||||:  
QY 422 -----TyrLeuValProThrVallysGlnGlyAsnTyr 432  
||| ||| ||||| :||||:  
Db 1782 TACCCTCTCTCTCTCTCTCTTACCTTAATAACTCTAACCAAACCTTT 1737  
||| ||| ||||| :||||:  
  
RESULT 10  
US-10-240-453/c  
; Sequence 25. Application US/10240453  
; GENERAL INFORMATION:  
; APPLICANT: OLER, Alexander  
; APPLICANT: PIEPERROCK, Christian  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA  
; TITLE OF INVENTION: Transcription  
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated  
; TITLE OF INVENTION: With DNA Transcription  
; FILE REFERENCE: 5013 1009  
; CURRENT APPLICATION NUMBER: US/10/240,453  
; CURRENT FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: PCT/EP01/03973  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 350  
; SEQ ID NO 25  
LENGTH: 5979  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
QY 10-240-453-25

Alignment Scores:  
Seq. Id. No.:  
Seq. Length:  
Identical Similarity:  
Percent Similarity:  
Gap Local Similarity:  
Mismatches:  
Indels:  
Gaps:

QY 09-910-186A-10 (1-450) x US-10-240-453-25 (1-5979)  
7 IlePheSerTyThrAsnAsnSerLeuLeuLysAsp-----IleIleAsnGlyTyrPhe 24  
|||||:|||||:|||||:|||||:  
Db 4240 ATATACACCATTTTCAACAATATTATTAAACAATAAACAATACGTAAATCAAAATTATCAA 4181  
|||||:|||||:|||||:|||||:  
QY 25 AsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAsp 44  
|||||:|||||:|||||:|||||:  
Db 4180 AAATATATCACAAAAACAATTTCTCTTAATATCAAAATCAAAATCAAAATCAAAATCAAAAT 4121  
|||||:|||||:|||||:|||||:  
QY 45 -ThrSerGlyTyrASnAlaGluValSerGluglu-----GlyAspVa 58  
|||||:|||||:|||||:|||||:  
Db 4120 TACTAATTTCCAAAAACAATATCTCCACAAACACTCCATTCAAAACAACATATTCACACAC 4061  
|||||:|||||:|||||:|||||:  
QY 58 lGlnLeuAsnProIle---PreProPhe-----AspPheLysLeuGlySerSergl 74  
|||||:|||||:|||||:|||||:  
Db 4060 CCAAAAAATCCAATCAACTTCTCTTAATTCACATATCTACCAAAACCACTATCTCTCCC 4001  
|||||:|||||:|||||:|||||:  
QY 74 yGluAspArgGlyLysValIleValThrGlnAsnGluSnIleValTyr----- 90  
|||||:|||||:|||||:|||||:



Copied from 10910186 on 05-05-2004



```

; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 107
; LENGTH: 6725
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-107
Alignment Scores:
  Score: 0.065 Length: 6725
  Percent Similarity: 105.50 Matches: 74
  Best Local Similarity: 37.32% Conservative: 54
  Query Match: 21.57% Mismatches: 116
  Indels: 100
  Gaps: 15
US-09-910-186A-10 (1-450) x US-10-240-485-107 (1-6725)
88 IleValTyrAsnSerMetTyrGluSerPheSerIleSerPheThrPheIleAsnLys 107
   ::::::::::::::::::::
4490 CTACTCTCAAACTCTTTTAT-----TACTATAACAATAATTTTATTATTAATCAAA 4437
   ::::::::::::::::::::
108 TrpValSerAsnLeuProGlyTyrThrIleGluAspSerValLysAsnSerGlyTrp 127
   ::::::::::::::::::::
4436 TAT-----TTTACTATTATAATAAAATTCATATAACAAC----- 4401
   ::::::::::::::::::::
128 SerIleGlyIleLeuSerAsnPheLeuValPhe----- 138
   ::::::::::::::::::::
4400 -----TTTACCNACTACTCTTTT-----TTTAAATAAAATCTACTCT 4353
   ::::::::::::::::::::
139 -----ThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
   ::::::::::::::::::::
4352 ATCAACAAATAAAATCAAT-----AACGCAATCTCAATTCATACACCTC 4305
   ::::::::::::::::::::
157 -----SerAsnAsnAlaProGlyTyrAsnLysTrpPheValThrValThr 172
   ::::::::::::::::::::
4304 TACTCTCCGGAATTCACAAATCTCTC-----ACCTCAACC 4269
   ::::::::::::::::::::
173 AsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLys 192
   ::::::::::::::::::::
4268 TCCCAATAAATAAACTACAAACATATACCAATACCAACCACTAATTTTATATTTTA 4209
   ::::::::::::::::::::
193 ValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIle 212
   ::::::::::::::::::::
4208 ATA-----AAACAATAATTCACCATATACCAAAATA 4176
   ::::::::::::::::::::
213 -----ProAspThrGlyLeuIleThrSerAspSerAsp 223
   ::::::::::::::::::::
4175 ATCTAAATCTCTTAACCTCGTAATCCGCGCTCGACCTCCAAATAATAAAATTAACA 4116
   ::::::::::::::::::::
224 AsnIleAsnMetTrpIleArgAspPheTyr-----IlePheAlaLysGluLeuAspGly 241
   ::::::::::::::::::::
4115 ATATAAAACACACACATCCGACCTACTTTTCCCACTACTTTTCCATATCTTACCTCC 4056
   ::::::::::::::::::::
242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrp 261
   ::::::::::::::::::::
4055 CTTTCCATATTAAACATATAAAATCATTTAAACCAACACATTTTAAAA----- 4005
   ::::::::::::::::::::
262 GlyAsnAspLeuArgTyrAsnLysGlnTyrTrpMetValAsnIleAspTyrLeuAsnArg 281
   ::::::::::::::::::::
4004 -----ATAAAACCTTACCACCAACATAT----- 3984
   ::::::::::::::::::::
282 TyrMetTyrAlaAsnSerArgGluIleValPheAsnThrArgAsnAsnAspPhe 301
   ::::::::::::::::::::

```

```

Db 3383 -----ATCCAACTAAATACTACCTTTTAAAAATAACAATCT- 3949
   ::::::::::::::::::::
Qy 302 AsnGluGlyTyrLysIleIleLeuLysArgGlyAsnThrAsnAspThrArgVal 321
   ::::::::::::::::::::
Db 3348 CAAAAAACTATACAACTTTTAAAAACAATATACCAAACTCAAAAAAACT----- 3895
   ::::::::::::::::::::
Qy 322 ArgGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341
   ::::::::::::::::::::
Db 3394 -----ACACTTCTATATCTACCTTAAAAAACAATCTTAACCCACACAAATATCA 3844
   ::::::::::::::::::::
Qy 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIle-----Tyr 358
   ::::::::::::::::::::
Db 3843 TCACCTTTATAATTACCTTTATTTTTCATCAATAAATCAAACTTAAATAATTACAAATTTAATTT 3784
   ::::::::::::::::::::
Qy 359 AlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGln 378
   ::::::::::::::::::::
Db 3783 TAATAACAATTCAAAAACCTTATTCAAAAATCAAAATAATCTCATTTTCAAACTAAA 3724
   ::::::::::::::::::::
Qy 379 Pro-----MetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLys 392
   ::::::::::::::::::::
Db 3723 CCTAATTTTAAATCTTCAATAATAAATAAAACCAATAACATAACTCATACCTATAATCC 3664
   ::::::::::::::::::::
Qy 393 SerAsnPhe 395
   ::::::::::::::::::::
Db 3663 CAACACTTT 3655

RESULT 15
US-10-240-454-49/c
; Sequence 49, Application US/10240454
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Metabolism
; FILE REFERENCE: 5013.1010
; CURRENT APPLICATION NUMBER: US/10/240,454
; PRIOR FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/04016
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 68
; SEQ ID NO 49
; LENGTH: 9265
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-454-49
Alignment Scores:
  Pred. No.: 0.116 Length: 9265
  Score: 105.00 Matches: 83
  Percent Similarity: 36.92% Conservative: 44
  Best Local Similarity: 24.13% Mismatches: 107
  Query Match: 4.39% Indels: 110
  DB: 6 Gaps: 17
US-09-910-186A-10 (1-450) x US-10-240-454-49 (1-9265)
Qy 133 SerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPhe 152
   ::::::::::::::::::::
Db 6345 TCAACTTTTACTATATTTTAAAAATCTCTA-----CTTTCCCAAAATTAATAAT 6292
   ::::::::::::::::::::
Qy 153 SerTyrAspIleSerAsnAlaProGlyTyrAsnLysTrpPheValThrValThr 172
   ::::::::::::::::::::

```

Db 6291 CACTACATTATTACGTCACAC-----ATAACACTCAATAT 6256  
 QY 173 AsnAsnMetMetGlyAsnMetLysLeu-TyrIleAsnGlyLysLeuIleAspThrIle 192  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 6255 AACAACTATTACTATTCAAAATAATATATACCAAGCACTATACAACTACTAA 6196  
 QY 192 sValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle----- 209  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 6195 AAAAACAACACTCCATACATTAATACAAATATCTATTAACCCATCGAATCCACTAAA 6136  
 QY 209 ----- 209  
 Db 6135 AACTATCCATACCAACAACATCCACAACCAACAAAAAACAACAAATTAACCTT 6076  
 QY 210 -----AsnLysIleProAspThrGlyLeuIle----- 218  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 6075 ACACAAACCTTATTATACAAAAAATACATCCCTTTACTTCCTCTCTATTATAATC 6016  
 219 -----ThrSerAspSerAspAsnIleAsnMetTrpIleAr 230  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 6015 CAAACCTTAACCTCTGCTACTCCACAACCTCTTCAAAATCACCTTAACCTC----- 5964  
 230 gAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn---IleLeuPheAs 249  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 5963 -AACTTCCCGTCAACAATTTCGAAACCAACCTCCCTTAATATTAACTACATATTATTCAC 5905  
 249 n-----Se 250  
 5904 ATTACTTCATATTATTTACCTCAATATATTTAAACGCACACTAAACCATATTTTA 5845  
 250 lleuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeu----- 265  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 5844 TTTAAATTTCAATATTCCTTCCAAATTTCTCCCTTTTAAACTACTATATTATAAT 5785  
 266 -ArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyr-A 285  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 5784 CCAATATCTTCCCATTTAC-----TTAAACATAACATTTTAAAT-----CACCC 5740  
 285 laAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGluGlyT 305  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 5739 CTAACTACTCAAAATAACACCAACCAATTAATCAATCAATCAATCAATCAATCAATCA 5680  
 305 yLys---IleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyG 324  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 5679 ATAACTATTCTCTATCCCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5627  
 324 lyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysA 344  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 5626 -----CTCCTTTTATCTCA---AATCTCATATCTATTAATCAATCAATCAATCAAT 5575  
 344 snGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArg 364  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 5574 ATAAACCGTC-----TCAACGTTTCTCAATTTAAITATATCTATTTACATAAC- 5522  
 364 luGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPromet----- 380  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 5521 -----TCCACAAATATCTCATACAAATATACAAATATATATATATATATATAT 5479  
 381 -----AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnG 397  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 5478 AACATTACCGAATAAT-----TATTTCCTCAACAAATAACTAAACAAATCTAAATA 5425  
 QY 397 lyGluAsn 399  
 ||||| : : : : :  
 Db 5424 TTCATAAT 5417

Search completed: November 7, 2002, 20:05:40  
 Job time : 62 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:41:56 ; Search time 43 Seconds

(without alignments)  
1006.058 Million cell updates/sec

Title: US-09-910-186a-10

Perfect score: 2391

Sequence: 1 MTIPNFISYTNNSLLKDI.....NYASLLESTHGWGVPVSE 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	2386	99.8	1291	2 S46431	botulinum neurotoxin
2	2386	99.8	1291	2 A49777	botulinum neurotoxin
3	1823	76.2	1285	2 S70582	botulinum neurotoxin
4	867.5	36.3	1276	2 S11435	botulinum neurotoxin
5	604.5	25.3	1297	2 S39791	botulinum neurotoxin - Clost
6	594.5	24.9	1252	2 S21178	botulinum neurotoxin
7	594	24.8	1251	2 JH0256	botulinum neurotoxin
8	593	24.8	1291	2 I40631	non-proteolytic bo
9	586.5	24.5	1296	1 BTCLAB	bontoxilysin (EC 3
10	581	24.3	1291	1 A48940	bontoxilysin (EC 3
11	569	23.8	1268	2 S33411	botulinum neurotoxin
12	554.5	23.2	1296	2 I40645	botulinum neurotoxin
13	549.5	23.0	1274	2 I40813	neurotoxin type F
14	425	17.8	1315	1 BTCLTN	tentoxilysin (EC 3
15	281.5	11.8	367	2 S48106	neurotoxin type E
16	279	11.7	366	2 S48110	neurotoxin type F
17	245.5	10.3	369	2 S48109	neurotoxin type F
18	230.5	9.6	1162	2 A47708	progenitor toxin n
19	230.5	9.6	1162	2 I40817	botulinum toxin no
20	213	8.9	1193	2 S68218	botulinum neurotoxin
21	199	8.3	1165	2 I40644	botulinum neurotoxin
22	199	8.3	1196	2 JQ1467	toxin, nontoxic co
23	199	8.3	1196	2 S46430	botulinum neurotoxin
24	191.5	8.0	1193	2 JQ4901	nontoxic-nonhemag
25	174.5	7.3	960	2 S72284	DNA-directed RNA p
26	171.5	7.2	1844	2 D71512	hypothetical prote
27	170.5	7.1	398	1 R3BYM1	ribosomal protein
28	168.5	7.0	1817	2 H71611	probable secreted
29	168	7.0	4688	2 F62885	hypothetical prote

ALIGNMENTS

RESULT 1

S46431

botulinum neurotoxin C1 - Clostridium botulinum phage 1C (strain C 468)

N.Alternate names: BoNT/C1 protein

C.Species: Clostridium botulinum phage 1C

A.Variety: strain C 468

C.Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 16-Jul-1999

C.Accession: S46431; S49107

R.Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.

Mol. Gen. Genet. 243, 631-640, 1994

A.Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxi

A.Reference number: S46426; MUID:94301293; PMID:8028579

A.Accession: S46431

A.Status: nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-1291 <HAU>

A.Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175

A.Experimental source: strain C 468

A.Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C.Superfamily: tetanus toxin

Query Match 99.8%; Score 2386; DB 2; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 1.7e-129;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTPNFISYTNNSLLKDIINEYFNINDSKILSCQNRKNTLVDTSGYNAEVSERGDYQLN 61

Db 843 TTPNFISYTNNSLLKDIINEYFNINDSKILSCQNRKNTLVDTSGYNAEVSERGDYQLN 902

QY 62 PTFPFDKLGSSGDRGVIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121

Db 903 PTFPFDKLGSSGDRGVIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962

QY 122 KNSGWSIGIISNLFVFLKQNEDESEQINSYDINSNAPGYNKWFVVTNNMGNNKI 181

Db 963 KNSGWSIGIISNLFVFLKQNEDESEQINSYDINSNAPGYNKWFVVTNNMGNNKI 1022

QY 182 YINGKLIDITKVKELTGINFSTITFEINKIPDGLITSDSDNINMIRDFYIFAKELDG 241

Db 1023 YINGKLIDITKVKELTGINFSTITFEINKIPDGLITSDSDNINMIRDFYIFAKELDG 1082

QY 242 KDINILFNSLQTNVVKDYGWNLDRYKEYMVDYLNRYWYANSQIVENTRRNNDF 301

Db 1083 KDINILFNSLQTNVVKDYGWNLDRYKEYMVDYLNRYWYANSQIVENTRRNNDF 1142

QY 302 NEGYKIIIRKRGNTNDRVRGDDILYFDMTINKKAYNLFMKNETMYADNHSSTEDIYAIG 361

Db 1143 NEGYKIIIRKRGNTNDRVRGDDILYFDMTINKKAYNLFMKNETMYADNHSSTEDIYAIG 1202

QY 362 LREQTKDINDNIIFQIQPNNTYYVYASQIFKSNFNGENISGICSGYTRFRLGGDWTRHN 421

Db 1203 LREQTKDINDNIIFQIQPNNTYYVYASQIFKSNFNGENISGICSGYTRFRLGGDWTRHN 1262

30 166.5 7.0 617 2 D96978  
31 159 6.6 2817 2 B97033  
32 157.5 6.6 762 2 E81371  
33 157 6.6 1465 2 A70199  
34 155.5 6.5 888 2 E82885  
35 155 6.5 1711 2 T18429  
36 155 6.5 3724 2 T18427  
37 154.5 6.5 2013 2 C71610  
38 153 6.4 1306 2 T28313  
39 150.5 6.3 1365 2 T18419  
40 150 6.3 437 2 T28180  
41 148 6.2 1127 2 T28317  
42 148 6.2 3973 2 B71612  
43 147.5 6.2 1436 2 D71618  
44 147.5 6.2 2485 1 H71621  
45 147 6.1 883 2 T18484

hypothetical prote  
uncharacterized pr  
probable periplasm  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
probable membrane  
ORF MSV152 probabl  
hypothetical prote  
hypothetical prote  
ORF MSV156 hypothe  
hypothetical prote  
hypothetical prote  
serine/threonine-s  
hypothetical prote



QY 422 YLVPVTKQGNVASYLLESTSTHWGFVPVSE 450  
|||||  
Db 1263 YLVPVTKQGNVASYLLESTSTHWGFVPVSE 1291

## RESULT 2

A49777  
botulinum neurotoxin type C1 precursor - Clostridium botulinum phage (type C, strain c-s)  
C:Species: Clostridium botulinum phage  
C:Date: 10-Mar-1994 #sequence\_revision 07-Apr-1994 #text\_change 23-Mar-2001  
C:Accession: S11291; A35396; S22166; A49777  
R:Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.;  
Nucleic Acids Res. 18, 4924, 1990  
A:Title: Nucleotide sequence of Clostridium botulinum C1 neurotoxin.  
A:Reference number: S11291; MUID:90370487; PMID:2204031

A:Accession: S11291

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-84, P' 86-1291 &lt;HAU&gt;

C:Cross-references: EMBL:X53751; NID:914905; PIDN:CAA37780.1; PID:914906

R:Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Takeshi, K.

Biochem. Biophys. Res. Commun. 171, 1304-1311, 1990

A:Title: The complete nucleotide sequence of the gene coding for botulinum type C-1 toxin

A:Reference number: A35396; MUID:91024998; PMID:2222445

A:Accession: A35396

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-869, R', 671-1291 &lt;TS1&gt;

R:Tsuzuki, K.; Kimura, K.; Fujii, N.; Yokosawa, N.; Oguma, K.

Submitted to the EMBL Data Library, December 1991

A:Description: Nucleotide sequence of the gene for one of the components of hemagglutinin

A:Reference number: S22163

A:Accession: S22166

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1291 &lt;TS2&gt;

C:Cross-references: EMBL:X62389; NID:9558175; PIDN:CAA44263.1; PID:940390

R:Kimura, K.; Fujii, N.; Tsuzuki, K.; Murakami, T.; Indoh, T.; Yokosawa, N.; Oguma, K.

Appl. Environ. Microbiol. 57, 1168-1172, 1991

A:Title: Cloning of the structural gene for Clostridium botulinum type C-1 toxin and who

A:Reference number: A49777; MUID:91282468; PMID:2059039

A:Accession: A49777

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-607 &lt;TS3&gt;

C:Cross-references: GB:D90210

Superfamily: tetanus toxin

Keywords: neurotoxin

Query Match 99.8%; Score 2386; DB 2; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 1.7e-129;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 TTPNFNFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61  
|||||  
Db 843 TTPNFNFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 902

QY 62 PIFPDFKLGSGEDRGKVIYVTQENIVYNSWESFSISFWIRINKWVSNLPGYTIIDSV 121  
|||||

Db 903 PIFPDFKLGSGEDRGKVIYVTQENIVYNSWESFSISFWIRINKWVSNLPGYTIIDSV 962

QY 122 KNSGWSIGIISNLFVTLKONEDSEQSINFSYDINNAPGYNKWFVTVTNMGMNMI 181  
|||||

Db 963 KNSGWSIGIISNLFVTLKONEDSEQSINFSYDINNAPGYNKWFVTVTNMGMNMI 1022

QY 182 YINGKLIDTIKVELTGINFSKTITFEINKIPDTGLTSDSDNINMWIRDFYIFAKELDG 241  
|||||

Db 1023 YINGKLIDTIKVELTGINFSKTITFEINKIPDTGLTSDSDNINMWIRDFYIFAKELDG 1082

QY 242 KDINILFNSLQYTNVVDYWGNDLYNKEYYWNIDYLNRYWYANSQIVFNTRRNNDF 301  
|||||

Db 1083 KDINILFNSLQYTNVVDYWGNDLYNKEYYWNIDYLNRYWYANSQIVFNTRRNNDF 1142

QY 302 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYATG 361  
|||||

Db 1143 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYATG 1202

QY 362 LREOTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISGICISGITYRFLGGDWYRHN 421  
|||||

Db 1203 LREOTKDINDNIIFQIQPMNNTYYASQIFKSNFNGENISGICISGITYRFLGGDWYRHN 1262

QY 422 YLVPVTKQGNVASYLLESTSTHWGFVPVSE 450  
|||||

Db 1263 YLVPVTKQGNVASYLLESTSTHWGFVPVSE 1291

## RESULT 3

S70582

botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sa  
C:Species: Clostridium botulinum phage d-sa

A:Note: host Clostridium botulinum type D (strain South Africa)

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Jun-2000

C:Accession: S70582

R:Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.

Biochim. Biophys. Acta 1307, 123-126, 1996

A:Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C

A:Reference number: S70582; MUID:96283801; PMID:8679691

A:Accession: S70582

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1285 &lt;MOR&gt;

C:Cross-references: EMBL:D38442; NID:91374775; PIDN:BA07477.1; PID:91374776

A:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit

a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic me

C:Superfamily: tetanus toxin

C:Keywords: disulfide bond; neurotoxin; transmembrane protein

F:1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted &lt;MAT1&gt;

F:448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted &lt;MAT2&gt;

Query Match 76.2%; Score 1823; DB 2; Length 1285;  
Best Local Similarity 77.7%; Pred. No. 3.8e-97;  
Matches 351; Conservative 37; Mismatches 56; Indels 8; Gaps 4;

QY 2 TTPNFNFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61  
|||||

Db 839 TTPNFNFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 898

QY 62 PIFPDFKLGSGEDRGKVIYVTQENIVYNSWESFSISFWIRINKWVSNLPGYTIIDSV 121  
|||||

Db 899 PIFPDFKLGSGEDRGKVIYVTQENIVYNSWESFSISFWIRINKWVSNLPGYTIIDSV 958

QY 122 KNSGWSIGIISNLFVTLKONEDSEQSINFSYDINNAPGYNKWFVTVTNMGMNMI 181  
|||||

Db 959 KNSGWSIGIISNLFVTLKONEDSEQSINFSYDINNAPGYNKWFVTVTNMGMNMI 1018

QY 182 YINGKLIDTIKVELTGINFSKTITFEINKIPDTGLTSDSDNINMWIRDFYIFAKELDG 241  
|||||

Db 1019 YINGKLIDTIKVELTGINFSKTITFEINKIPDTGLTSDSDNINMWIRDFYIFAKELDG 1078

QY 242 KDINILFNSLQYTNVVDYWGNDLYNKEYYWNIDYLNRYWYANSQIVFNTRRNNDF 301  
|||||

Db 1079 KDINILFNSLQYTNVVDYWGNDLYNKEYYWNIDYLNRYWYANSQIVFNTRRNNDF 1138

QY 302 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 360  
|||||

Db 1139 NEGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 1193

QY 361 GUREQTKD-INDNIIFQIQPMNNTYYASQIFKSNFNGENISGICISGITYRFLGGDWY 418  
|||||

Db 1194 GALDQPMDEIRKYSFIQPCNTFDYASQLFSSNATNRLGILSIGSYSLGDDYWF 1253

QY 419 RNYLVPTVKQGNVASYLLESTSTHWGFVPVSE 450  
|||||

Db 1254 RNYLVPTVKQGNVASYLLESTSTHWGFVPVSE 1285

RESULT 4  
S11455  
botulinum neurotoxin type D - Clostridium botulinum  
C:Species: Clostridium botulinum  
C:Date: 18-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1999  
C:Accession: S11455  
R:Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegl  
Nucleic Acids Res. 18, 5556, 1990  
A:Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type  
A:Reference number: S11455; MUID:91016853; PMID:2216736  
A:Accession: S11455  
A:Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1276 <BIN>  
A:Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match	36.3%;	Score 867.5;	DB 2;	Length 1276;
Best Local Similarity	43.2%;	Pred. No. 3e-42;		
Matches 195;	Conservative 83;	Mismatches 146;	Indels 27;	Gaps 11;

[illegible]

RESULT 5  
S39791  
neurotoxin - Clostridium botulinum  
C:Species: Clostridium botulinum  
C>Date: 07-Oct-1994 #sequence\_revision 01-Dec-1995 #text\_change 16-Jul-1999  
C:Accession: S39791  
R:Campbell, K.; Collins, M.D.; East, A.K.  
Biochem. Biophys. Acta 1216, 487-491, 1993  
A:Title: Nucleotide sequence of the gene coding for Clostridium botulinum ((  
A:Reference number: S39791; MUID:94092745; PMID:8268233  
A:Accession: S39791  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1397 <CAM>  
A:Cross-references: EMBL:X74162; NID:9441275; PIDN:CAA52275.1; PID:g441276  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

[illegible]

RESULT 6

botulinum neurotoxin type E precursor - Clostridium botulinum  
S21178

C:Species: Clostridium botulinum  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Oct-1999  
C:Accession: S21178; S48107; JH0257; B35294; A60027; S18111  
R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.  
Eur. J. Biochem. 204, 657-667, 1992

A:Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin  
A:Reference number: S21178; MUID:92174922; PMID:1541280

A:Accession: S21178

A:Molecule type: DNA

A:Residues: 1-1252 <WHE>

A:Cross-references: EMBL:X52683; NID:540397; PIDN:CAA44558.1; PID:g40398

R:Campbell, K.D.; Collins, M.D.; East, A.K.  
J. Clin. Microbiol. 31, 2255-2262, 1993

A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific  
A:Reference number: S48103; MUID:94013372; PMID:8408542

A:Accession: S48107

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 616-982 <CAM>

A:Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, January 1993

R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.  
Biochem. Biophys. Res. Commun. 183, 107-113, 1992

A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum t  
A:Reference number: JH0256; MUID:92181428; PMID:1543481

A:Accession: JH0257

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA





C:Genetics: A:Gene: atx; bota  
C:Function: A:Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associated  
C:Superfamily: tetanus toxin  
C:Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane protein  
F:2-444/product: bontoxilysin A light chain #status experimental <LIGHT>  
F:445-1296/product: bontoxilysin A heavy chain #status experimental <HVT>  
F:223,227/binding site: zinc (His) #status predicted  
F:224/Active site: Glu #status predicted

Query Match 24.5%; Score 586.5; DB 1; Length 1296;  
Best Local Similarity 31.6%; Pred. No. 4.4e-26;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Qy 3 IPFNIFSYTNLSKLDIINEYFNNDKILSKLQNKNTLVDTSGYNAEVSSEGDVQLNP 62  
849 IPFQSLKYVDNQRLSTFTFYIKNTISLIRYENHLIDLSRYASKINIGSKVNDP 908  
63 IFPFD--FKLGSSGDEGRGVIVTQENIYVNSMYESFISFWIRINKVSNLP--GYT 116  
909 IDKQIQIQLFNLESS--KIEVILKNAIYVNSMYENFSTFWIRIPKYPNSISLNNEYT 963  
117 IIDSVKNSNGSIGIISNVLFTLKQNEDEQSINFSYDINSNAPGY-NKWPFFVTVNNM 175  
964 IINCWENSGKSLNGEITWLTQTEIKQRVFKYSQMINISDYINRWTFVITNNR 1023  
176 MGNKIYKINGKLDTIKVKELTGINSKTTFTFEIKKIPDTGLTSDNINWIRDFYIF 235  
1024 LNNKTYINGLIDDKPSINLGHASNNIMFKDGRDT-----HYIWIYKFNLF 1075  
236 AKELGDKDINTLFSLOYTVNVKDYNGDLRYNKYKMYNIDYLNRYMANS----- 287  
1076 DKELNEKEIKLDYQNSGLKDFWGDYLDQKPYMLNLYDPNKYVDVNVGIRGYM 1135  
288 ---RQIVFTR-RNNDFNEGYKIIKIRGTNTNTRVGGDILYFDMTINKAYNLFM 342  
1136 LKGRPGSVMTNLYNSLSLXGKTKFIKKYASGNKNINVRNDRVYINVVKNKEYRL-- 1193  
343 KNETMYADNHSDEIYVIGREQTQKNDNMIIFQIOPMNTYVYASQIFKSNFNGENTSG 402  
1194 -----ATNASQAGVERKLSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNGND 1246  
403 ICSTGTRFRGLGGWYRHNYLVPTVROGNTVASLLETS-----THWGFVPVSE 450  
1247 IGFITGFHQF-----NNIAKLVASWYNRQIERSRSLTGCSEWEIFPVD 1289

RESULT 10  
A:8940  
Bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum  
Alternative names: botulinum neurotoxin type B (BoNT/B)  
Species: Clostridium botulinum  
Date: 19-Dec-1993 #sequence, revision 18-Nov-1994 #text\_change 18-Jun-1999  
Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574  
R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.  
Appl. Environ. Microbiol. 58, 2345-2354, 1992  
A:Title: Molecular cloning of the Clostridium botulinum structural gene encoding the type B neurotoxin  
A:Reference number: A48940; MUID:92384550; PMID:1514783  
A:Accession: A48940  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1291 <HVE>  
A:Cross-references: GB:M81186; NID:g144734; PIDN:AAA3211.1; PID:g144735  
A:Experimental source: type B, Danish  
A:Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBP:112081); this publication  
R:Campbell, K.D.; Collins, M.B.; East, A.K.  
J. Clin. Microbiol. 31, 2255-2262, 1993  
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific isoforms  
A:Reference number: S48103; MUID:94013372; PMID:8408542  
A:Accession: S48105  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 634-994 <CAM>  
A:Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783  
A:Experimental source: proteolytic type B, strain NCTC 7273  
R:Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.  
submitted to the EMBL Data Library, April 1992  
A:Description: Partial amino acid sequence of botulinum neurotoxin type B and compared with the sequence of botulinum neurotoxin type A  
A:Reference number: S21575  
A:Accession: S21575  
A:Molecule type: DNA  
A:Residues: 36-217, 'G', '219-224', 'S', '226-246' <SZA>  
A:Cross-references: EMBL:211934; NID:g40383; PIDN:CAA77991.1; PID:g40384  
R:Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars  
J. Biol. Chem. 267, 14721-14729, 1992  
A:Title: Minimal essential domains specifying toxicity of the light chains of tetanus toxin  
A:Reference number: A42871; MUID:92340509; PMID:1634516  
A:Accession: A42871  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-313, 'S', '315-451' <KUR>  
A:Experimental source: strain Okra  
A:Note: sequence extracted from NCBI backbone (NCBIP:109365)  
R:Dasgupta, B.R.; Datta, A.  
Biochimie 70, 811-817, 1988  
A:Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with botulinum neurotoxin type A  
A:Reference number: S07155; MUID:89000987; PMID:3139097  
A:Accession: S07155  
A:Molecule type: protein  
A:Residues: 2-29, 'M', '31-45' <DAS>  
A:Accession: S08562  
A:Molecule type: protein  
A:Residues: 442-463, 'R', '465-467' <DA2>  
R:Schmidt, J.J.; Sathyanarayanan, V.; Dasgupta, B.R.  
Arch. Biochem. Biophys. 238, 544-548, 1985  
A:Title: Partial amino acid sequences of botulinum neurotoxins types B and E.  
A:Reference number: S07128; MUID:85197963; PMID:3888113  
A:Accession: S07128  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 2-16 <SCH1>  
A:Accession: S08573  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 2-17 <SCH2>  
A:Accession: S08574  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 442-459 <SCH3>  
R:Schiaivo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, N.  
Nature 359, 832-835, 1992  
A:Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic modification of synaptobrevin  
A:Reference number: S2125; MUID:93063293; PMID:1331807  
A:Contents: annotation  
C:Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synapses  
C:Genetics: A:Gene: bont/b  
C:Function: A:Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2  
C:Superfamily: tetanus toxin  
C:Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc  
F:2-441/product: bontoxilysin B light chain #status experimental <LIGHT>  
F:442-1291/product: bontoxilysin B heavy chain #status experimental <HVT>  
F:230,234/binding site: zinc (His) #status predicted  
F:231/Active site: Glu #status predicted

Query Match 24.3%; Score 581; DB 1; Length 1291;  
Best Local Similarity 32.9%; Pred. No. 9e-26;  
Matches 161; Conservative 82; Mismatches 168; Indels 78; Gaps 21;

Qy 3 IPFNIFSYTNLSKLDIINEYFNNDKILSKLQNKNTLVDTSGYNAEVSSEGDVQLNP 62  
849 IPFQSLKYVDNQRLSTFTFYIKNTISLIRYENHLIDLSRYASKINIGSKVNDP 908  
63 IFPFD--FKLGSSGDEGRGVIVTQENIYVNSMYESFISFWIRINKVSNLP--GYT 116  
909 IDKQIQIQLFNLESS--KIEVILKNAIYVNSMYENFSTFWIRIPKYPNSISLNNEYT 963  
117 IIDSVKNSNGSIGIISNVLFTLKQNEDEQSINFSYDINSNAPGY-NKWPFFVTVNNM 175  
964 IINCWENSGKSLNGEITWLTQTEIKQRVFKYSQMINISDYINRWTFVITNNR 1023  
176 MGNKIYKINGKLDTIKVKELTGINSKTTFTFEIKKIPDTGLTSDNINWIRDFYIF 235  
1024 LNNKTYINGLIDDKPSINLGHASNNIMFKDGRDT-----HYIWIYKFNLF 1075  
236 AKELGDKDINTLFSLOYTVNVKDYNGDLRYNKYKMYNIDYLNRYMANS----- 287  
1076 DKELNEKEIKLDYQNSGLKDFWGDYLDQKPYMLNLYDPNKYVDVNVGIRGYM 1135  
288 ---RQIVFTR-RNNDFNEGYKIIKIRGTNTNTRVGGDILYFDMTINKAYNLFM 342  
1136 LKGRPGSVMTNLYNSLSLXGKTKFIKKYASGNKNINVRNDRVYINVVKNKEYRL-- 1193  
343 KNETMYADNHSDEIYVIGREQTQKNDNMIIFQIOPMNTYVYASQIFKSNFNGENTSG 402  
1194 -----ATNASQAGVERKLSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNGND 1246  
403 ICSTGTRFRGLGGWYRHNYLVPTVROGNTVASLLETS-----THWGFVPVSE 450  
1247 IGFITGFHQF-----NNIAKLVASWYNRQIERSRSLTGCSEWEIFPVD 1289

Db 896 ---KNQFKLTSSA---NSKLRVTONONIFNSVFLDPSVFWIRIPKYANDGIQNIHNEYT 951  
QY 117 IIDSVKNSGWSIGILISNPLVFTLKQNEDESQSINFSYDISNNAPGY-NKWFVTVYNNM 175  
Db 952 IINCNNKNSGWSIGIRGNRIITWIDINGTKSVFFSEYINREDSISEINRWFVTTNN- 1010  
QY 176 MGNMAYIYINGKLIDITKVKELTGTFNFKSTIFFEINKPIDTGLTSDSDNIN-MWIRDFYI 234  
Db 1011 LNNAKIYINGKLESNTDKIDREVIANGEIIFKLD-----GDIDRQPIWMKVFSEI 1061  
QY 235 FAKELDGDNDINFLNSLOYTNVADYWGNDLRYNKKEYMYNIDLYNRYMYANS-----RQI 290  
Db 1062 FNTLSQSNIEERYKIQSYSEYLDKDFWGNPLMYNKKEYMFNAGKNKSYIKLKDKSPVGEI 1121  
QY 291 VENTRRNN-----DNFEGYKIIIRIGNT---NDTRVGGDILYFD-MTINKN--- 336  
1122 LTRSYNQNSKYINVRDYIYGEKIIFIR-KNSQSINDDIVRKEDYIYLDFFNLNWEVRV 1180  
337 -AYNLFMKNETMADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYIASQ-IFKSN 394  
1181 YNYKFKKEE-----EKLFLAPISDSDEFYN---TIQKEYDEQPTYSQQLFKK- 1227  
395 FNGENISGICSGTGYRFRLLG-----DWYRHNYLVPTVKOGNYASLLESTST 441  
1228 -DEESTDELIGLIGIRHYESGIVFEYKDYFCISKWY-----LKEVRKPYNLKL---GC 1278  
442 HMGFVPVSE 450  
1279 NMQFIPKDE 1287  
RESULT 11  
3411  
botulinum neurotoxin type F - Clostridium barati  
C:Species: Clostridium barati  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Jul-1999  
C:Accession: S33411; S31860  
R:Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.;  
R:Microbiol. Lett. 108, 175-182, 1993  
A:Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin  
A:Reference number: S33411; MUID:9325228; PMID:8486245  
A:Accession: S33411  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1268 <THO>  
C:Cross-references: EMBL:X68262; NID:949138; PIDN:CAA48329.1; PID:g49139  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin  
Query Match 23.8%; Score 569; DB 2; Length 1268;  
Best Local Similarity 32.7%; Pred. No. 4.3e-25;  
Matches 154; Conservative 82; Mismatches 167; Indels 68; Gaps 17;  
2 TPFNFSTYNNSLKLDINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGQVQLN 61  
Db 832 SIFPELSEYNTDKILHILIRFYKRDSSILNKKYENNRIDSGYSGNISNGDIYII 891  
QY 62 PIPFDFKLGSGEDRGKVIYTONENIVYNSWIESFSISFWIRINKW--VSNLPG-YTII 118  
Db 892 STNRNCFGYISS--RLSEVNTQNTNIYNSYQNFVSFWVRIPKYNNLKNLNNETII 949  
QY 119 DSVK-NNSGWSIGITISNPLVFTLKQNEDESQSINFSY-----DISNAPGYNKWFVTVN 173  
Db 950 NCMRRNNSGWSISLNNIITWIDTQTNQKLVFNVTQMDISDY---INKWTFVITN 1006  
QY 174 NMGNMAYIYINGKLIDITKVKELTGTFNFKSTIFFEINKPIDTGLTSDSDNINMWIRDFY 233  
Db 1007 NRIGHSKLTINGLDQKISILNGLNIHVDDNLLFKIVGCNDTRYV-----GIRIFK 1057  
QY 234 IFAKELDGDNDINFLNSLOYTNVADYWGNDLRYNKKEYMYNIDLYNRYMYANSRQIVFN 293  
Db 1058 IFNMELDKTEIETLVHSEPDSTILKDFWGNLYLNKYYILLNLLKPNMSVTKNSDILNIN 1117

QY 294 TRR-----NNDNFEGYKIIIRIGNTN---DTRVGGDILYFDMTINKKAYNL 340  
Db 1118 KORGLYSTNTIFSNARLYTGVVEIIRKV-GSTDTSTNDFNVRKNDTVIINVDGNSIQL 1176  
QY 341 FMKNETMADNHSTEDIYAIGLREQTKDINDNIIFQIQPMNNTYYIASQIFKSNFNGEN 399  
Db 1177 -----YADVSTSAVEKTIKLRISNSNTNSN--QMIIMDSIGDNDCTMNFXTN-NGND 1225  
QY 400 ISGICSGITGYRFRLL-----GGDWYRHNYLVPTVKOGNYASLLESTSTHWGFV 446  
Db 1226 -----IGLLGFHLNLYVASSWYKYNIRNTRNNGCF-----WSFI 1260  
RESULT 12  
140645  
botulinum neurotoxin type A - Clostridium botulinum  
C:Species: Clostridium botulinum  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 16-Jul-1999  
C:Accession: I40645  
R:Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.  
Res. Microbiol. 144, 547-556, 1993  
A:Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type  
A:Reference number: I40645; MUID:94143603; PMID:8310180  
A:Accession: I40645  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1296 <RES>  
C:Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin  
Query Match 23.28; Score 554.5; DB 2; Length 1296;  
Best Local Similarity 30.8%; Pred. No. 3e-24;  
Matches 147; Conservative 81; Mismatches 183; Indels 67; Gaps 12;  
QY 3 IPFNFTSTNLSLKLDINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGQVQLNP 62  
Db 849 IPFQLSKYVDNKKLLSTTEYIKNVTLSILSYVKKDDLDLSRYGAKINIGDRVYDS 908  
QY 63 IPFPDFKLGSGEDRGKVIYTONENIVYNSWIESFSISFWIRINKWVSNL---PGYTIID 119  
Db 909 IDKNOIKLINLESSTIEVIL--KNAIVNSMYENFSTFWIKIPKFSKINLNNETIIN 966  
QY 120 SVKNSGWSIGITISNPLVFTLKQNEDESQSINFSYDISNNAPGY-NKWFVTVYNNMGN 178  
Db 967 CIENNSGKVSINLYGEIITWIDQNKONIORVVKYSQMVNLSQVINEWIFVTITNNRLTK 1026  
QY 179 MKIYINGKLIDITKVKELTGTFNFKSTIFFEINKPIDTGLTSDSDNINMWIRDFYIFAKE 236  
Db 1027 SKIYINGRLIDQKIPISNLGN-HASNKIMFKLDGCRD-----PRYIMKYFNLPDKE 1078  
QY 239 LDGKDINTLFNSLOYTNVADYWGNDLRYNKKEYMYNIDLYNRYMYANS----- 287  
Db 1079 LNEKEIKLDYSQNSGILKDFWGNLYQYDPYVMLNLPKPKYVDVNNIGIRGYMYLKG 1138  
QY 288 --RQIVFTRNNNDNFNGYKIIIRIGNTNDRVYRGGDILYFDMTINKKAYNLFMKNE 345  
Db 1139 PRGSVVTNIVLNSTLYEGTKFIKKVAGSNEEDIVRNDRVYINVVYKKEKYL----- 1193  
QY 346 TMYADNHSTEDIYAIGLREQTKDINDNIIFQI-----QPMNNTYYIASQIFKSNFNG 397  
Db 1194 ---ATNASQAGVEKILSALEIPDVGN--LSQVYVYKSKDDQGIKRNK-----CKMNLQD 1241  
QY 398 ENISGICSGITGYRFR-----RLGGDWYRHNYLVPTVKOGNYASLLESTSTHWGFVPSVSE 450  
Db 1242 NNGNDIGFIFGHLYDNIKLVASNWN-----RQVGRAS--RTFGCSWEFIPVDD 1289  
RESULT 13  
140813  
botulinum neurotoxin type F - Clostridium botulinum  
C:Species: Clostridium botulinum

Copied from 109 10 186 of 05-05-2004

Query Match

Query Match	17.8%;	Score 425;	DB 1;	Length 1315;
Best Local Similarity	26.4%;	Pred. No. 8.5e-17;		
Matches 134: Conservative	90;	Mismatches 169;	Indels 114;	Gaps 21;

3 IPF----NIFYTNNSLLKDIINEYFNINDSKILSLQNRKNTLVDGSGYNAEVSSEGDV 58  
||| | : : | | : | ||| : : ||| : |  
859 IPFSYSKNIDCWVNDEEDIDVT-----LKKTSLKLDINDIISDLGSFGNSSVITYPDA 912

59 QLNPIFPDFKLGSGE-----DRGKVIVTQENIVNSMYEFSISFWIRNK-WYS 110

```
Db 913 QIVP-----GINKAHLVNNESSEVIVHKAMDIEYNDMFNNFTVFWLRVVKVSAS 964
QY 111 NLP-----GYTIIDSVKNN-----SGMSIGIISNLFVFTLKQNEDESQSINFSDISNNA 160
Db 965 HLEOYGTNEYSIISMKKHSLSIGSGNSVSLKGNLIWTLKDSAGEVRQITFR-DLPDKF 1023
QY 161 PCY--NKWFFVTNMMGNMKIYINGKLDTIKVKEITGTFNFKITTFEINKIPDTGLI 218
Db 1024 NAYLANKWFFITINDRUSSANLVINGVWGSAEITGLGAIRDENNITLKDR----- 1076
QY 219 TSDSDNINMWIRDEYIFAKEIDGKDINILFNSQYTNVVKDYMGNDLRNKEYIMYNI-- 276
Db 1077 -CNNNQVSDIKPRIFCKALNPKIEKLYSLITFLRDFWGNPLRYDTEYLLIPVAS 1135
QY 277 -----DYL---NRYMYANSROIVFNTRNNNDNFEGYKIIIKIRGNTN-DTRV 321
Db 1136 SKQVQKNIIDYMYLTNAPSYTNGKLNIIYRRLYN-----GLKFIKRYTPNNEIDSEV 1190
QY 322 RGGDILYFDMTINKAY-----NLFMKNETMYADNHSTEDIYAIGLREOTKINDNI 373
Db 1191 KSGDEIKLVSYNNNEHIVGYPKDGNAFNLDRLRVGINAPGIPLYKKMEAVK----- 1244
QY 374 IFQIQPMNTYYIASQIFKSNFNGENISGICIGTYRFRLGDD-----WYRHNLYL 423
Db 1245 -----LRDLKTYSVQL--KLYDDKNAS-LGLVGTHTNGQIGNDPNRDILIASNMY-FNHL 1294
QY 424 VPTVKQGNVASLLESTHTHWGFPVSE 450
Db 1295 KDKI-----LGCDWYFVPIDE 1310
RESULT 15
S48106
neurotoxin type E - Clostridium botulinum (fragment)
Species: Clostridium botulinum
Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
Accession: S48106
Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2253-2262, 1993
Title: Gene probes for identification of the botulinum neurotoxin gene and specific id
Reference number: S48103; MUID:94013372; PMID:8408542
Accession: S48106
Status: preliminary; nucleic acid sequence not shown; translation not shown
Molecule type: DNA
Residues: 1-367 <CAM>
Cross-references: EMBL:X70818; NID:9407784; PIDN:CAA50149.1; PID:9407785
Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
Superfamily: tetanus toxin
Keywords: neurotoxin
Query Match 11.8%; Score 281.5; DB 2; Length 367;
Best Local Similarity 35.8%; Pred.No.3e-09;
Matches 58; Conservative 41; Mismatches 56; Indels 7; Gaps 3;
QY 2 TIPFNIFSYTNNSLLKDIINEYFNINDSKILSLQNRKNTLVDTSGYNAEYSEGDVQLN 61
Db 207 SIFFKLSSYTDKILSYFNKFFKIKSSVNLNMYKNDKYVDTSGYDSNININGDYKY 266
QY 62 PIFFPDKLGSSGEDRGKVIYTONENIVNMYESFSISFWIRI-----NKWVSNLPCYTI 117
Db 267 PTNKNQF--GIYNDKLSEYNSQNDYIIYDNKYNFISFWIRIPNYDNDKIVNYYNEITI 324
QY 118 IDSVK--NNSGWSIGIISNLFVFTLKQNEDESQSINFSDISN 158
Db 325 INCMRDNNSGWSVNLNHNELIWLQDNAGINQKLAFNIGNAN 366
```

Search completed: November 7, 2002, 14:46:49  
Job time : 53 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 13:46:50 ; Search time 24 seconds  
(without alignments)  
777.681 Million cell updates/sec

Title: US-09-910-186a-10  
Perfect score: 2391  
Sequence: 1 MTIPFNFSYTNNSLLKDI.....NYASLLESTHGWGFPVSE 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 112892 seqs, 41476328 residues 112892  
Total number of hits satisfying chosen parameters:  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2386	99.8	1290	1	BXCL_CLOBO
2	867.5	36.3	1276	1	BXD_CLOBO
3	604.5	25.3	1296	1	BXG_CLOBO
4	594	24.8	1250	1	BXE_CLOBO
5	586.5	24.5	1295	1	BXAL_CLOBO
6	583	24.4	1250	1	BXE_CLOBO
7	581	24.3	1290	1	BXE_CLOBO
8	554.5	23.2	1295	1	BXA2_CLOBO
9	549.5	23.0	1274	1	BXF_CLOBO
10	425	17.8	1314	1	TETX_CLOTE
11	230.5	9.6	1162	1	BXEN_CLOBO
12	230.5	9.6	1162	1	BXEN_CLOBO
13	199	8.3	1196	1	BXCN_CLOBO
14	164.5	6.9	396	1	RYAR_YEAST
15	155	6.5	1138	1	CTAB_BACUK
16	147	6.1	1076	1	RCOB_ASTLIO
17	144.5	6.0	2136	1	YCF2_MARPO
18	142.5	6.0	3135	1	S230_PLAFO
19	140.5	5.9	1024	1	RCOB_PLAFO
20	138.5	5.8	339	1	RYAR_CANGA
21	138.5	5.8	476	1	RC3_SACBA
22	138.5	5.8	2366	1	TOXB_CLODI
23	138	5.8	692	1	Y650_METUA
24	136.5	5.7	537	1	ARP_PLAFA
25	135.5	5.7	760	1	ETF2_MSEPV
26	135	5.6	559	1	ENS2_YEAST
27	134	5.6	1138	1	C7AA_BACTU
28	133.5	5.6	655	1	YKDA_WYCCA
29	132.5	5.5	1104	1	COLA_CLOPE
30	131.5	5.5	451	1	ARP2_PLAFA
31	131.5	5.5	575	1	RCPC_PLAFA
32	131.5	5.5	828	1	PMFC_PROMI
33	131.5	5.5	987	1	YD94_METUA

RESULT 1									
ID	BXCL_CLOBO	STANDARD;	PRT;	1250	AA.				
AC	P18640;								
DT	01-NOV-1990	(Rel. 16, Created)							
DT	01-NOV-1990	(Rel. 16, Last sequence update)							
DT	15-JUN-2002	(Rel. 41, Last annotation update)							
DE	Botulinum neurotoxin type C1 precursor (EC 3.4.24.69) (BoNT/C1)								
DE	(Bontoxilysin Cl):								
OS	Clostridium botulinum.								
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;								
OC	Clostridium.								
OX	NCBI_TaxID=1491;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=90370487; PubMed=2204031;								
RA	Hauser D., Eklund M.W., Kurazona H., Binz T., Niemann H., Gill D.M.,								
RA	Boquet P., Popoff M.R.;								
RT	"Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";								
RL	Nucleic Acids Res. 18:4924-4924(1990).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=Type C Stockholm / C-ST;								
RX	MEDLINE=91024998; PubMed=2222445;								
RA	Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,								
RA	Yokosawa N., Takeshi K., Syuto B., Oguma K.;								
RT	"The complete nucleotide sequence of the gene coding for botulinum								
RT	type C1 toxin in the C-ST phage genome.";								
RL	Biochem. Biophys. Res. Commun. 171:1304-1311(1990).								
RN	[3]								
RP	SEQUENCE OF 2-25.								
RC	STRAIN=Type C Stockholm / C-ST;								
RX	MEDLINE=88153072; PubMed=2450068;								
RA	Tsuzuki K., Yokosawa N., Syuto B., Ohishi I., Fujii N., Kimura K.,								
RA	Oguma K.;								
RT	"Establishment of a monoclonal antibody recognizing an antigenic site								
RT	common to Clostridium botulinum type B, C1, D, and E toxins and								
RT	tetanus toxin.";								
RL	Infect. Immun. 56:898-902(1988).								
RN	[4]								
RP	IDENTIFICATION OF SUBSTRATE.								
RX	MEDLINE=94038966; PubMed=7901002;								
RA	Blasi J., Chapman E.R., Yamasaki S., Binz T., Niemann H., Jahn R.;								
RA	"Botulinum neurotoxin C1 blocks neurotransmitter release by means of								
RT	cleaving HPC-1/syntaxin.";								
RL	EMBO J. 12:4821-4828(1993).								
CC	-!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER								
CC	RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED								
CC	AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD								
CC	WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT								
CC	INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC								
CC	ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.								
CC	-!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the								
CC	neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No								
CC	detected action on small molecule substrates.								
CC	-!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A								

P47549 mycoplasma  
Q58313 methanococc  
Q91562 fowlpox vir  
P45355 haemophilus  
P58145 astasia lon  
P49331 streptococc  
P45385 haemophilus  
P27625 plasmodium  
Q9q8k4 myxoma viru  
P14248 plasmodium  
P03879 saccharomyc  
Q9pral ureaplasma

ALIGNMENTS

CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

CC -1- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C  
 CC STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE  
 CC BACTERIOPHAGE.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X66433; CAA47060.1; -

CC EMBL; X72793; CAA51313.1; -

CC EMBL; X53751; CAA37780.1; -

CC EMBL; D90210; BAA14235.1; -

CC EMBL; X62389; CAA44263.1; -

CC PIR; S11291; S11291.

CC PIR; A35396; A35396.

CC PIR; A43503; A43503.

CC HSSP; P10845; 3BTA.

CC MEROPS; M27.002; -

CC InterPro; IPR000395; Bontoxilysin.

CC InterPro; IPR000130; Zn\_Mpopeptidase.

CC Pfam; PF01742; Peptidase\_M27; 1.

CC PRINTS; PR00760; BONTOXILYSIN.

CC PRODOM; PD001963; Bontoxilysin; 1.

CC PROSITE; PS00142; ZINC\_PROTEASE; 1.

CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.

CC INIT\_MET 0 0

CC CHAIN 1 448

CC METAL 449 1290

CC ACT\_SITE 228 228

CC METAL 229 229

CC METAL 232 232

CC DISULFID 436 452

CC CONFLICT 84 84

CC P -> T (IN REF. 2).

CC SEQUENCE 1290 AA; 148734 MW; 71FBE379F9712956 CRC64;

CC Query Match 99.8%; Score 2386; DB 1; Length 1290;

CC Best Local Similarity 100.0%; Pred. NO. 8.2e-135;

CC Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC 2 TIPNIFSYTNLSKDIINEYFNINDSKILSLQNKNTLVDTSGYNAEVSSEGVDQLN 61

CC 842 TIPNIFSYTNLSKDIINEYFNINDSKILSLQNKNTLVDTSGYNAEVSSEGVDQLN 901

CC 62 PIFPDPFKLGGSGEDRGKVIYQTNENIVNYSMESEISFIRINKWNSLPGYTIIDSV 121

CC 902 PIFPDPFKLGGSGEDRGKVIYQTNENIVNYSMESEISFIRINKWNSLPGYTIIDSV 961

CC 122 KNNCSWSTGILSNFLVFLTKQNESEOSINFSVDISNNAPGYKNKFFVTVNNMGNMKI 181

CC 962 KNNCSWSTGILSNFLVFLTKQNESEOSINFSVDISNNAPGYKNKFFVTVNNMGNMKI 1021

CC 182 YNGKLIDTIKVELGTINFSKTTTFEINKIPDTGLTSDSDNNINMIRDFYFAKELDG 241

CC 1022 YNGKLIDTIKVELGTINFSKTTTFEINKIPDTGLTSDSDNNINMIRDFYFAKELDG 1081

CC 242 KDINILFNSLQTVNVDYWGNDLRYNKEYMYNIDYLNRYMANSQIVFNTFRNNDF 301

CC 1082 KDINILFNSLQTVNVDYWGNDLRYNKEYMYNIDYLNRYMANSQIVFNTFRNNDF 1141

CC 302 NRGYKIIIRKRGNDNRVRGGDILYFDMTINKKAYNLFMKMETMYADNHSTEDYIAG 361

CC 1142 NRGYKIIIRKRGNDNRVRGGDILYFDMTINKKAYNLFMKMETMYADNHSTEDYIAG 1201

QY 362 LRQTKDINDNIIFOQPMNNTYYASQLEKSNFNGENISGICSTGYRFRLLGGDWYRHN 421

DB 1202 LRQTKDINDNIIFOQPMNNTYYASQLEKSNFNGENISGICSTGYRFRLLGGDWYRHN 1261

QY 422 YLPTVKQGNYSALLESTSTHMGFVPVSE 450

DB 1262 YLPTVKQGNYSALLESTSTHMGFVPVSE 1290

# RESULT 2

ID BMD\_CLOBO

AC PI9321; STANDARD; PRT; 1276 AA.

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BONT/D)

DE (Bontoxilysin D).

GN BOTD.

OS Clostridium botulinum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium

OX NCBL\_TaxID=1491;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BVD/-3;

RX MEDLINE=91016853; PubMed=2216736;

RA Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,

RA Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;

RA "Nucleotide sequence of the gene encoding Clostridium botulinum

RA neurotoxin type D";

RL Nucleic Acids Res. 18:5556-5556(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CB16;

RX MEDLINE=93042276; PubMed=1420572;

RA Sunagawa H., Ohyanaka T., Watanabe T., Inoue K.;

RA "The complete amino acid sequence of the Clostridium botulinum type D

RA neurotoxin, deduced by nucleotide sequence analysis of the encoding

RA phage d-16 phi genome";

RL J. Vet. Med. Sci. 54:905-913(1992).

RN [3]

RP PARTIAL SEQUENCE.

RC STRAIN=D-SA, and D-1873;

RX MEDLINE=89339741; PubMed=2668193;

RA Morishita K., Syuto B., Kubo S., Oguma K.;

RA "Molecular diversity of neurotoxins from Clostridium botulinum type D

RA strains";

RL Infect. Immun. 57:2885-2891(1989).

RN [4]

RP IDENTIFICATION OF SUBSTRATE.

RX MEDLINE=94230352; PubMed=8175689;

RA Yamasaki S., Baumeister A., Binz T., Blas J., Link E., Cornille F.;

RA Roques B., Fyke E.M., Suedhof T.C., Jahn R., Niemann H.;

RA "Cleavage of members of the synaptobrevin/VAMP family by types D and

RA F botulinum neurotoxins and tetanus toxin";

RL J. Biol. Chem. 269:12764-12772(1994).

CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER

CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED

CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD

CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT

CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC

CC ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF

CC SYNAPTOSOMES-1 AND -2.

CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the

CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No

CC detected action on small molecule substrates.

CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A

CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,

CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL

CC FORMATION AND TOXIN BINDING, RESPECTIVELY.

CC -1- SUBCELLULAR LOCATION: Secreted.



```

FT DISULFID 435 449 INTERCHAIN (PROBABLE).
SQ SEQUENCE 1296 AA; 149013 MW; DC8E47E15F665C31 CRC64;

Query Match
Best Local Similarity 32.0%; Score 604.5; DB 1; Length 1296;
Matches 156; Conservative 87; Mismatches 172; Indels 73; Gaps 16;

Qy 2 TIPNIFSYNNSLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVEEGDVLN 61
Db 839 SIPPDLSTYKDTLIQVFNYSINSSNAILSLYRGRLDSSGYGATMVGSDVFN 898
Qy 62 PIFPFDKLGSSGDRGKVTQONENIVNSMYESFSISFWIRI-----LPGY 115
Db 899 DIGNQKFLNNS--ENSNIHTAHSQKFFVYDSMFDNFSINFWRTPKYNNNDIQTLYQNEY 956
Qy 116 TIISVKNNSGWSGIISNLFVFLTKONEDSEOSINFSYDISNAPCY-NKWFVTVNN 174
Db 957 TIISCIKNDGKWSKISGRIIWLTDVNAKSKSIFFSIKDNISDYINKWFSTIIND 1016
Qy 175 MGNMKIYINGKLDITKVKELGFINFSKTTTEINKIPDTGLTSDSDNINMWIRDFYI 234
Db 1017 RLGNANIYINGSLKSKKILNLRINSNDIDFKLINCTDTKF-----VWIKDFNI 1068
Qy 235 FAKELGKDNILFNSLQYNNVYKWDGNDLRYNKVYVWINDYLNRYVYANSRQIVFNT 294
Db 1069 FGRLEATEVSSLYWIOSSTNTLKFQWGNPLRDYQYLFNQGMQNIYKYFSKASGET 1128
Qy 295 RRNNDFNE-----GKIIIRI-----RGNTNDTRVGGDILYFDMTINNKAYNLF 341
Db 1129 APRTN-FNNAINYONLYGLRFTIKKASNRINNDIVREGDYIYLI----- 1177
Qy 342 MKNETYADNHTEDIYAGLRQTKDNDNIIFQIQWNN--TYVYASQI---PKSNF 395
Db 1178 -----DNISDES-YRVYVLSNKEIQQLF--LAPINDPTFDVQIKYKYEKTYI 1226
Qy 396 NGENIS-----GICSGTGYRFLGSDWYRH-NYLVTYKQNGYASILESTS-----TH 442
Db 1227 NCQILCEKDKTFLGFGIGKFKVYDGYVWTDYDNF--CISQWYLRRISGNINKRLGNC 1284
Qy 443 WGFVPVSE 450
Db 1285 WQFIPVDE 1292

SEQUENT 4
BXE_CLOBU STANDARD; PRT; 1250 AA.
P30995;
01-JUL-1993 (Rel. 26, Created)
01-JUL-1993 (Rel. 26, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
(Bontoxilysin E).
Clostridium butyricum.
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
NCBI_TaxID=1492;
[1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 43181, and ATCC 43755;
RX MEDLINE=92181428; PubMed=1543481;
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT "Sequences of the botulinum neurotoxin E derived from Clostridium
BT botulinum type E (strain Beluga) and Clostridium butyricum (strains
ATCC 43181 and ATCC 43755).";
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).
[2]
SEQUENCE OF 1-251 FROM N.A.
RC STRAIN=BL6340;
RX MEDLINE=91237316; PubMed=2033376;
RA Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA Yokosawa N., Yashiki T., Oguma K.;
RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum

```

```

type E toxin gene from Clostridium butyricum strain BL6340.";
J. Gen. Microbiol. 137:519-525(1991).
[3]
SEQUENCE OF 1-48.
RC STRAIN=3262;
RA Gimenez J., Foley J., Dasgupta B.R.;
RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT partial sequence and comparison.";
RL FASEB J. 2:A1750-A1750(1988).
CC -I- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -I- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO
CC detected action on small molecule substrates.
CC -I- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X62088; CAA43998.1; -
CC EMBL; X53180; CAA37321.1; -
CC PIR; JH0256; JH0256.
CC PIR; S16145; S16145.
CC HSP; P10845; 38TA.
CC MEROPS; M27.002; -
CC
CC InterPro; IPR000395; Bontoxilysin.
CC InterPro; IPR000130; Zn_Mtpeptidse.
CC Pfam; PF01742; Peptidase_M27; 1.
CC PRINTS; PR00760; BONTOXILYSIN.
CC ProDom; PD001963; Bontoxilysin; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
KW INIT_MET 0 0
FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT METAL 211 211 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 212 212 BY SIMILARITY.
FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 411 425 INTERCHAIN (PROBABLE).
FT CONFLICT 229 229 K->M (IN REF 2).
SQ SEQUENCE 1250 AA; 143265 MW; 6171B5B2C312857 CRC64;

Query Match
Best Local Similarity 32.8%; Score 594; DB 1; Length 1250;
Matches 150; Conservative 90; Mismatches 160; Indels 66; Gaps 17;

Qy 2 TIPNIFSYNNSLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVEEGDVLN 61
Db 821 SIPPDLSTYKDTLIQVFNYSINSSNAILSLYRGRLDSSGYGATMVGSDVFN 880
Qy 62 PIFPFDKLGSSGDRGKVTQONENIVNSMYESFSISFWIRI-----NKWVSNLPGVTI 117
Db 881 PTNKNQF--GLYNDKLESEVNISSNDYIYDNKYKFSISFWIRPNYDNKIVNVNNEVTI 938
Qy 118 IDSVK--NNSGWSGIISNLFVFLTKONEDSEOSINFSYDISNAPCY-NKWFVTVNNK 175
Db 939 INCRDNNNSGKWSKSLNHNHETIWTLDQNSGINQKLAFLAFNYGNANGISDYINKWIFVTITNDR 998

```

176 MGNMXYINGKLDITKIKVLTGKINFSKTTITEINKIPDTGLTSDSNINNNWIRDFYIF 235  
 999 LGOSKUYINGNLDDKSIILNIGHVSDNLEKTYNCSTYRI-----GIRYFNIF 1049  
 236 AKELDKGKINILFNSLOYTNVVKYNGNDLRNKEYYMWNI-----DYLNRYWYANSROIV 291  
 1050 DKELDETEIQTLYNNEPNANILKDEWNGYLLDYKEYLLNVLKPNFNFR--RUDSTLSI 1107  
 292 FNTFRN--NNDNEGYKIIKRI--RGNTNDRVRGGDILYFDMINNKAYNLFKMKNETM 347  
 1108 NNIRSTILLANRYSIGIKVQIWNSSNDNLVRKNDVOYINF-VASKTHLL-----PL 1161  
 348 YADNHSPEIYAGLEQRTQKINDNTIIFOQPMNNTYYVASOIFKSNNGENISICSIG 407  
 1162 YADTATTNKETIKISSGNRFNQVVM-----NSVGNCTMNFKN--NGNN-----IG 1208  
 408 TYFRP-----LGGOWYRNYLVPVKOGYASLLESTSH--WGFW 446  
 1209 LLGPKADTVVASTWY-----YTHMRDNINSNGFFWNI 1241

RESULT 5  
 BVAL\_CLOBO STANDARD; PRT; 1295 AA.  
 P10845; P18639; P01561;  
 01-JUL-1989 (Rel. 11, Created)  
 01-JUL-1993 (Rel. 26, Last annotation update)  
 15-JUN-2002 (Rel. 41, Last annotation update)  
 Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)  
 (Bontolysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-  
 chain; Botulinum neurotoxin A, heavy chain].  
 BOTA OR BNA OR ATX.  
 Clostridium botulinum.  
 Clostridia; Clostridiales; Clostridiaceae;  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.  
 NCBI\_TaxID=1491;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=NTC 2916;  
 MEDLINE=90235866; PubMed=2185020;  
 Thompson D.E., Brem J.K., Oultram J.D., Swinfield T.-J.,  
 Shone C.C., Atkinson T., Melling J., Minton N.P.;  
 "The complete amino acid sequence of the Clostridium botulinum type A  
 neurotoxin, deduced by nucleotide sequence analysis of the encoding  
 gene";  
 Eur. J. Biochem. 189:73-81(1990).  
 [2]  
 SEQUENCE FROM N.A.  
 STRAIN=62A;  
 MEDLINE=90264400; PubMed=2160960;  
 Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;  
 "The complete sequence of botulinum neurotoxin type A and comparison  
 with other clostridial neurotoxins";  
 J. Biol. Chem. 265:9153-9158(1990).  
 [3]  
 SEQUENCE OF 1-65 FROM N.A.  
 STRAIN=62A;  
 MEDLINE=97016817; PubMed=8863443;  
 East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
 "Organization and phylogenetic interrelationships of genes encoding  
 components of the botulinum toxin complex in proteolytic Clostridium  
 botulinum types A, B, and F: evidence of chimeric sequences in the  
 gene encoding the nontoxic nonhemagglutinin component";  
 Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
 [4]  
 SEQUENCE OF 1-34 FROM N.A.  
 STRAIN=Hall;  
 MEDLINE=89350959; PubMed=2669749;  
 Betley M.J., Somers E., Dasgupta B.R.;  
 "Characterization of botulinum type A neurotoxin gene: delineation of  
 the N-terminal encoding region";  
 Biochem. Biophys. Res. Commun. 162:1388-1395(1989).  
 [5]  
 SEQUENCE OF 1-18 FROM N.A.  
 STRAIN=Type A NIH;  
 MEDLINE=96096783; PubMed=8521962;  
 Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;  
 "Molecular characterization of two forms of nontoxic-nonhemagglutinin  
 components of Clostridium botulinum type A progenitor toxins";  
 FEBS Lett. 376:41-44(1995).  
 [6]  
 SEQUENCE OF 1-16.  
 MEDLINE=84178501; PubMed=6370252;  
 Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;  
 "Partial amino acid sequence of the heavy and light chains of  
 botulinum neurotoxin type A";  
 Biochem. Biophys. Res. Commun. 119:900-904(1984).  
 [7]  
 SEQUENCE OF 1-46.  
 Dasgupta B.R., Foley J., Niece R.;  
 "Partial sequence of the light chain of botulinum neurotoxin type A";  
 Biochemistry 26:4162-4162(1987).  
 [8]  
 SEQUENCE OF 1-5 AND 444-456.  
 MEDLINE=91120847; PubMed=2126206;  
 Dasgupta B.R., Dekleva M.L.;  
 "Botulinum neurotoxin type A: sequence of amino acids at the  
 N-terminus and around the nicking site";  
 Biochimie 72:661-664(1990).  
 [9]  
 SEQUENCE OF 448-454 AND 872-895.  
 MEDLINE=89024662; PubMed=3178218;  
 Sartymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;  
 "Botulinum neurotoxin type A: cleavage of the heavy chain into two  
 halves and their partial sequences";  
 Arch. Biochem. Biophys. 266:142-151(1988).  
 [10]  
 SEQUENCE OF 448-482.  
 MEDLINE=85285016; PubMed=3896784;  
 Shone C.C., Hambleton P., Melling J.;  
 "Inactivation of Clostridium botulinum type A neurotoxin by trypsin  
 and purification of two tryptic fragments. Proteolytic action near  
 the COOH-terminus of the heavy subunit destroys toxin-binding  
 activity";  
 Eur. J. Biochem. 151:75-82(1985).  
 [11]  
 IDENTIFICATION OF SUBSTRATE.  
 MEDLINE=94063091; PubMed=8243676;  
 Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,  
 Benfenati F., Wilson M.C., Montecucco C.;  
 "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
 COOH-terminal peptide bonds";  
 FEBS Lett. 335:99-103(1993).  
 [12]  
 IDENTIFICATION OF SUBSTRATE.  
 MEDLINE=94124495; PubMed=8294407;  
 Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
 Jahn R., Niemann H.;  
 "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins";  
 J. Biol. Chem. 269:1617-1620(1994).  
 [13]  
 MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.  
 MEDLINE=21556941; PubMed=11700044;  
 Rigoni M., Caccini F., Johnson E.A., Montecucco C., Rossetto O.;  
 "Site-directed mutagenesis identifies active-site residues of the  
 light chain of botulinum neurotoxin type a";  
 Biochem. Biophys. Res. Commun. 288:1231-1237(2001).  
 [14]  
 X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).  
 MEDLINE=98455071; PubMed=9783750;  
 Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;  
 "Crystal structure of botulinum neurotoxin type A and implications  
 for toxicity";  
 Nat. Struct. Biol. 5:898-902(1998).  
 CC  
 !- FUNCTION: Inhibits acetylcholine release. The botulinum toxin

binds with high affinity to peripheral neuronal presynaptic membrane, is then internalized by receptor-mediated endocytosis. The C-terminus of the heavy chain (H) is responsible for the adherence of the toxin to the cell surface while the N-terminus mediates transport of the light chain from the endocytic vesicle to the cytosol. After translocation, the light chain (L) hydrolyzes the 197-Gln-Arg-198 bond in SNAP-25, thereby blocking neurotransmitter release. Inhibition of acetylcholine release results in flaccid paralysis, with frequent heart or respiratory failure.

!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.

!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a heavy chain (H).

!- SUBCELLULAR LOCATION: Secreted.

!- PHARMACEUTICAL: Available under the name BOTOX (Allergan) for the treatment of strabismus and blepharospasm associated with dystonia and cervical dystonia. Also used for the treatment of hemifacial spasm and a number of other neurological disorders characterized by abnormal muscle contraction.

!- MISCELLANEOUS: There are seven antigenically distinct forms of botulinum neurotoxin: Types A, B, C1, D, E, F, and G.

!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

!- DATABASE: NAME-BOTOX product information Web site; WWW="http://www.botox.com/index.jsp?hps&productinfo".

!- DATABASE: NAME-Protein Spotlight;

NOTE-Issue 19 of February 2002;

WWW="http://www.expaasy.org/spotlight/articles/spt1019.html".

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; X52066; CAA36289.1; -  
EMBL; M30196; AAR23262.1; -  
EMBL; X92973; CAA63551.1; -  
EMBL; D67030; BAAL1051.1; -  
EMBL; M27892; AAR23269.1; -  
PIR; A35294; BTCLAB.  
PIR; S09492; S09492.  
PDB; 3BTA; 01-OCT-99.  
MEROPS; M27.002; -.  
InterPro; IPR000395; Bontoxilysin.  
InterPro; IPR000130; Zn\_Mtpeptidse.  
Pfam; PF01742; Peptidase\_M27; 1.  
PRINTS; PR00760; BONTOKILYSIN.  
ProDom; PD001963; Bontoxilysin; 1.  
PROSITE; PS00142; ZINC\_PROTEASE; 1.  
Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;  
Pharmaceutical; 3D-structure.  
INIT\_MET 0 0  
CHAIN 1 447 BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.  
CHAIN 448 1295 BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.  
METAL 222 222 ZINC (CATALYTIC).  
ACT\_SITE 223 223  
METAL 226 226 ZINC (CATALYTIC).  
METAL 261 261 ZINC (CATALYTIC).  
DISULFID 429 453 INTERCHAIN.  
DISULFID 1234 1279  
TRANSMEM 626 646 POTENTIAL.  
TRANSMEM 655 675 POTENTIAL.  
VARIANT 26 26 V -> A.  
MUTAGEN 261 261 E->A: DRASTIC DECREASE IN ENZYMACTIC ACTIVITY.  
MUTAGEN 265 265 F->A: DECREASE IN ENZYMACTIC ACTIVITY.  
MUTAGEN 365 365 Y->A: DECREASE IN ENZYMACTIC ACTIVITY.  
CONFLICT 1 1 P -> Q (IN REF. 1).  
CONFLICT 479 479 E -> P (IN REF. 9).

FT CONFLICT 875 875 T -> L (IN REF. 8).  
FT CONFLICT 891 891 S -> K (IN REF. 8).  
SQ SEQUENCE 1295 AA; 149322 MW; 858342F754862579 CRC64;  
  
Query Match 24.5%; Score 586.5; DB 1; Length 1295;  
Best Local Similarity 31.6%; Pred.No. 1.1e-27;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;  
  
QY 3 IPPNIFSYTNNSLLKDIINIEYFNINNDKSLISLQNRKNTLVDTSGYNAEYSEGDVOLNP 62  
DB 848 IPFQLSKYVDNORLLSTFTTEYIKNIINTSILNRYESNHLIDLSRVASKINIGSKVNFDP 907  
  
QY 63 IPPPD---FKLGSSGDEGRGVIVTQENIYVNSYSEFSISFWIRINKWVSNLP---GYT 116  
DB 908 IDKNOIQIENFESS-----KIEVLKNAIVNSMYENFSTFWIRIPKFNFSISLNNEYT 962  
  
QY 117 IIDSVKNNKSGWISGIISNLFVTLKQNEDESEQISNFSDISNNAPY-NKWFVFTVNNM 175  
DB 963 IINCENNSGKWSLYNGEIIWTLQDQEIQRVVFYSQMINISDYINRWIEFTVITNR 1022  
  
QY 176 MGNMKIYINGKLIDTIKVKELTGINFSTTTFEINKPTDPTGLITSDDNINMKIRDFYIF 235  
DB 1023 LNNKIIYINGKLIDQKIPISNIGNIHANNIMFKLDGCRDT-----HRYIKWIFNLF 1074  
  
QY 236 AKELDKDINILFNSLOYNVVKDYWDNLRNKEYYVWVNDIYLNRYMYANS-----287  
DB 1075 DKELNEKEIKDLYDNQNSGILKDFWDGYLDYQDKPYMLNLDPNKYVDVNVYVGIQRYMY 1134  
  
QY 288 ---RQIVFNTR-RNNNDFENEGYKIIIRKIRGTNDPRVRGGDILYFDMTINNKNLPM 342  
DB 1135 LKPGRGVYNTIYLNSSLYRGTFPIKKVASKNKNINRNDRVINNVYVKKETL-- 1192  
  
QY 343 KNETYADNHSTEDIAIAGLRQTKDINDNIIFIQPMNNTIYASQIFKSNFNGENISG 402  
DB 1193 -----ATNASQAGVEKILSALEIPDVGNLSQVYVYMKSKNDQGITNKC-KWNLQDNNGND 1245  
  
QY 403 ICSIGTYFRGLGGWYRNLYLVVVKOGNTASILESTS-----THWGFVPVSE 450  
DB 1246 IGFIFGHOF-----NNIAKLIVASWYRNQRISRSRTLGCSWEFIPVDD 1288  
  
RESULT 6  
BXE\_CLOBO STANDARD; PRT; 1250 AA.  
ID BXE\_CLOBO  
AC Q00496;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (Bont/E)  
DE (Bontoxilysin E).  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Beluga.  
RX MEDLINE=92181428; PubMed=1543481;  
RA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;  
RT "Sequences of the botulinum neurotoxin E derived from Clostridium botulinum type E (strain Beluga) and Clostridium butyricum (strains ATCC 43181 and ATCC 43755).";  
RT Biochem. Biophys. Res. Commun. 183:107-113(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92174922; PubMed=1541280;  
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;  
RT "The complete amino acid sequence of the Clostridium botulinum type E neurotoxin, derived by nucleotide-sequence analysis of the encoding gene.";  
RL Eur. J. Biochem. 204:657-667(1992).  
RN [3]  
RP SEQUENCE OF 1-251 FROM N.A.

MEDLINE=90264400; PubMed=2160950;  
Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;  
"The complete sequence of botulinum neurotoxin type A and comparison  
with other clostridial neurotoxins.";  
J. Biol. Chem. 265:9153-9158(1990).  
[4]  
SEQUENCE OF 1-13.  
MEDLINE=95197963; PubMed=3888113;  
Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;  
"Partial amino acid sequences of botulinum neurotoxins types B and  
E.";  
Arch. Biochem. Biophys. 238:544-548(1985).  
[5]  
SEQUENCE OF 419-426.  
MEDLINE=90344918; PubMed=2116911;  
Gimenez J.A., Dasgupta B.R.;  
"Botulinum neurotoxin type E fragmented with endoproteinase Lys-C  
reveals the site trypsin nicks and homology with tetanus  
neurotoxin.";  
Biochimie 72:213-217(1990).  
[6]  
IDENTIFICATION OF SUBSTRATE.  
MEDLINE=94063091; PubMed=8243676;  
Schiavoni G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,  
Benfante F., Wilson M.C., Montecucco C.;  
"Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
COOH-terminal peptide bonds.";  
FEBS Lett. 335:99-103(1993).  
[7]  
IDENTIFICATION OF SUBSTRATE.  
MEDLINE=94124495; PubMed=8294407;  
Binz T., Blaszi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
Jahn R., Niemann H.;  
"Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";  
J. Biol. Chem. 269:1617-1620(1994).  
-!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-  
181 BOND IN SNAP-25.  
-!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
neuroexocytosis apparatus, synaptobrevin, SNAP25 or syntaxin. No  
detected action on small molecule substrates.  
-!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
FORMATION AND TOXIN BINDING, RESPECTIVELY.  
-!- SUBCELLULAR LOCATION: Secreted.  
-!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X62089; CAA43999.1; -;  
EMBL; X62683; CAA44558.1; -;  
PIR; A60027; A60027.  
PIR; B35294; B35294.  
PIR; JH0257; JH0257.  
PIR; S08575; S08575.  
PIR; S16111; S16111.  
PIR; S21178; S21178.  
HSP; P10845; 3ETA.  
MEROPS; M27.002; -;  
InterPro; IPR000395; Bontoxilysin.

RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92384550; PubMed=1514783;  
 RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,  
 RA Minton N.P.;  
 RT "Molecular cloning of the Clostridium botulinum structural gene  
 RT encoding the type B neurotoxin and determination of its entire  
 RT nucleotide sequence.";  
 RL Appl. Environ. Microbiol. 58:2345-2354(1992).  
 RN [2]  
 RP SEQUENCE OF 35-245 FROM N.A.  
 RC STRAIN=NCTC 7273;  
 RA Szabo B.A., Pemberton J.M., Desmarchelier P.M.;  
 RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 633-993 FROM N.A.  
 RC STRAIN=NCTC 7273;  
 RA MEDLINE=94013372; PubMed=8408542;  
 RA Campbell K., East A.K., Collins M.D.;  
 RT "Gene probes for identification of the botulinum neurotoxin gene and  
 RT specific identification of neurotoxin types B, E, and F.";  
 RL J. Clin. Microbiol. 31:2255-2262(1993).  
 RN [4]  
 RP SEQUENCE OF 1-44 AND 441-466.  
 RC STRAIN=657;  
 RA MEDLINE=89000987; PubMed=3139097;  
 RA Dasgupta B.R., Datta A.;  
 RT "Botulinum neurotoxin type B (strain 657): partial sequence and  
 RT similarity with tetanus toxin.";  
 RL Biochimie 70:811-817(1988).  
 RN [5]  
 RP SEQUENCE OF 1-16 AND 441-458.  
 RC STRAIN=OKRA;  
 RA MEDLINE=85197963; PubMed=3886113;  
 RA Schmidt J.J., Sathyanarayanan V., Dasgupta B.R.;  
 RT "Partial amino acid sequences of botulinum neurotoxins types B and  
 RT E.";  
 RL Arch. Biochem. Biophys. 238:544-548(1985).  
 RN [6]  
 RP IDENTIFICATION AS ZINC-PROTEASE.  
 RA MEDLINE=93054694; PubMed=1429690;  
 RA Schlavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;  
 RT "Botulinum neurotoxins are zinc proteases.";  
 RL J. Biol. Chem. 267:23479-23483(1992).  
 RN [7]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RA MEDLINE=93063293; PubMed=1331807;  
 RA Schlavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,  
 RA Dasgupta B.R., Montecucco C.;  
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 RT by proteolytic cleavage of synaptobrevin.";  
 RL Nature 359:832-835(1992).  
 RN [8]  
 RP FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF  
 SYNAPTOSOMALIN-2.  
 RN [9]  
 RP CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 detected action on small molecule substrates.  
 RN [10]  
 RP SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIANE CHANNEL  
 FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 RN [11]  
 RP SUBCELLULAR LOCATION: Secreted.  
 RN [12]  
 RP MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 RN [13]  
 RP SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 RN [14]  
 RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M81186; AAA23211.1; -;  
 DR EMBL; Z11934; CAA77991.1; -;  
 DR EMBL; X70817; CAA50148.1; -;  
 DR PIR; S07128; S07128;  
 DR PIR; S07155; S07155;  
 DR PIR; S08562; S08562;  
 DR PIR; S08573; S08573;  
 DR PIR; S08574; S08574;  
 DR PIR; A48940; A48940;  
 DR HSSP; P10845; 3BTA;  
 DR MEROPS; M27.002; -;  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR000130; Zn\_Mtpeptidse.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 KW INIT\_MET 0  
 FT CHAIN 1 440 BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.  
 FT CHAIN 441 1290 BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.  
 FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 230 230 BY SIMILARITY.  
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 436 445 INTERCHAIN (PROBABLE).  
 FT CONFLICT 29 29 T -> M (IN REF. 4).  
 FT CONFLICT 217 217 R -> G (IN REF. 2).  
 FT CONFLICT 224 224 A -> S (IN REF. 2).  
 FT CONFLICT 463 463 S -> R (IN REF. 4).  
 SQ SEQUENCE 1290 AA; 150670 MW; D21746E2C024DF43 CRC64;  
 Query Match 24.3%; Score 581; DB 1; Length 1290;  
 Best Local Similarity 32.9%; Pred No. 2.3e-27;  
 Matches 161; Conservative 82; Mismatches 168; Indels 78; Gaps 21;  
 QY 3 IPFNFTSYNNLSLKIDINEYNNINDSKLSQNRKNTLVDTSGYNAEYSEGVQLNP 62  
 DB 835 MPFOLSIYNTDILIEFMFNKYNSEILNLIILNRYKDNLDLSGYGAKVYDGVGLND 894  
 QY 63 IPPDFKLGSSEGRGVIVTONENIVYNSVSEFSISFWIRINKW---VSNL--PGYT 116  
 DB 895 --KNQFKLTSA--NSKIRVTQNNILFNSVFLDFSVFWIRIPKYNKDCIQNYIENET 950  
 QY 117 IIDSYKNNSGWSIGIISNLFVTLKQNEDEQSEINFSYDISNNAPGY-NKWFVTVYNNM 175  
 DB 951 IINCNNKNSGWSKISIRGNRIITWLDINGKTKSVFEYINREDISEYINRWFVYITNN- 1009  
 QY 176 MGNMKIYNGKLIDITIKVKELTGINFSKITFEINKTPDGLITSDSNIN-MWIRDFVI 234  
 DB 1010 LNNAKIYNGKLESTWDIKDIREVANGELIIFKLD-----GIDRTQFIWMKYSFI 1060  
 QY 235 FAKELDGKIDINLFNSLQYTNVVKDYNGNLDNRKNKEYVMYNDILNRYMYANS----RQI 290  
 DB 1061 FNTELSQSNIEERYKIQSYSEYKDFWGNPLMYKNYMFENAGNKNKYIKLKDKSPWGEI 1120  
 QY 291 VFNFRNNN-----DFNEGYKIIIKRIGNT---NDTRVGRGDILYFD-WTINKK--- 336  
 DB 1131 ITRSKYNSKYNRYRDLYIGEKFIIR-XXNSQSINDDIVRKEDYIYDLFFNLNQEWRV 1179  
 QY 337 -AYNLFMKNETMYADNHSTEDIYAIGLRETKDINDNIFQIQPMNNYTYASQ-IFKSN 394  
 DB 1180 YTYXFKKEE-----EKLFLAPISDSDFYN---TIQKEYDEQPYSQLFKK- 1226  
 QY 395 FNGENISICIGYRFLRG-----DWVRHNYLVPTVKQGNYSALLSESTST 441  
 DB 1227 -DEESTDEIGLIGHREYVESGIVPEYKDYFCISKWY-----LKEVKKRPYNLKL---GC 1277



QY 442 HNGFVPVSE 450  
 Db 1278 NWQFIPKDE 1286

RESULT 8  
 BXA2\_CLOBO STANDARD; PRT; 1295 AA.  
 AC Q45894; P77780;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)  
 DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-chain; Botulinum neurotoxin A, heavy-chain].  
 DE BOTA OR BNA OR ATX.  
 DE Clostridium botulinum.  
 DE Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 DE Clostridium.  
 DE NCBI\_TaxID=1491;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=Kyoto-F;  
 MEDLINE=94143603; PubMed=8310180;  
 Willens A., East A.K., Lawson P.A., Collins M.D.;  
 "Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A associated with infant botulism: comparison with other clostridial neurotoxins.";  
 Res. Microbiol. 144:547-556(1993).  
 [2]  
 SEQUENCE OF 1-65 FROM N.A.  
 STRAIN=Kyoto-F;  
 MEDLINE=97016817; PubMed=8863443;  
 East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
 "Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic Clostridium botulinum types A, B, and F: evidence of chimeric sequences in the gene encoding the nontoxic nonhemagglutinin component.";  
 Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
 [-] FUNCTION: Inhibits acetylcholine release. The botulinum toxin binds with high affinity to peripheral neuronal presynaptic membrane, is then internalized by receptor-mediated endocytosis. The C-terminus of the heavy chain (H) is responsible for the adherence of the toxin to the cell surface while the N-terminus mediates transport of the light chain from the endocytic vesicle to the cytosol. After translocation, the light chain (L) hydrolyzes the 197-Gln-1-Alg-198 bond in SNAP-25, thereby blocking neurotransmitter release. Inhibition of acetylcholine release results in flaccid paralysis, with frequent heart or respiratory failure (By similarity).  
 [-] CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.  
 [-] SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a heavy chain (H) (By similarity).  
 [-] SUBCELLULAR LOCATION: Secreted.  
 [-] MISCELLANEOUS: There are seven antigenically distinct forms of botulinum neurotoxin: Types A, B, C1, D, E, F, and G.  
 [-] SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 -----  
 THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X73423; CAA51824.1; -;  
 EMBL; X87974; CAA61234.1; -;  
 HSP; P10845; BETA.  
 MROPS; M27.002; -;  
 InterPro; IPR000395; Bontoxilysin.

DR InterPro: IPR000130; Zn\_MTPeptidse.  
 DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; Hydrolase; Metalloprotease; Zinc.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 FT INIT\_MET 0 0  
 FT CHAIN 1 447 BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.  
 FT CHAIN 448 1295 BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.  
 FT METAL 222 222 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 223 223 BY SIMILARITY.  
 FT METAL 226 226 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 429 453 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 1234 1279 BY SIMILARITY.  
 FT TRANSMEM 626 646 POTENTIAL.  
 FT TRANSMEM 655 675 POTENTIAL.  
 SQ SEQUENCE 1295 AA; 149279 MW; 5DA04A13D98D6372 CRC64;

Query Match 23.2%; Score 554.5; DB 1; Length 1295;  
 Best Local Similarity 30.8%; Pred. No. 8.6e-26;  
 Matches 147; Conservative 81; Mismatches 183; Indels 67; Gaps 12;

QY 3 IPFNIPSYTNNLLKDIINEYFNNINDSKILSLQNKNTLVDTSGYNAEVSSEGQVQLNP 62  
 Db 848 IPQLSKYVDNKKLLSTFTTEYIKNIYNTSILSVYKDDLDLSRYGAKINIGDRVYDS 907  
 QY 63 IFPPDFKLSSGSDRGKIVTQENIVNYSMEYSPSISFWIRINKWSNL---PGYTIID 119  
 Db 908 IDKNQIKLINESSTIEVIL--KNAIVNYSMEYSPSISFWIRINKWSNL---PGYTIID 119  
 QY 120 SVKNKSGSIGISNLFVTLKONEDSEOSINFSYDYSNNAPGY-NKAFVFTVTNNMGN 178  
 Db 966 CIENNSCKVSLNYGEIITLQDNKONIQRVFKSQVMYNSIDYINRWLFTVITNNRLTK 1025  
 QY 179 MKIYINGKLDTIKVKELTGINSFTTTEINKIPDTGLTSDSNINNMWIDFYFAKE 238  
 Db 1026 SKIYINGRLIDQKPSINLGNHASNKIMFKDCRD-----PRRYIMIKYFNLFQKE 1077  
 QY 239 LDGKDINLNSLQYINVYKDYNGDLRYNKYVMYNDILNRYMANS----- 287  
 Db 1078 LNKETKDLSDSNGSGLKDFWNGYLDKPYMLNLFDPKRYVDVNNIGRGYWLKG 1137  
 QY 288 --RQIVENTRRNNDFNEGKIIIRIGTNTNTRVGGDILYFDMTINNKAYNLFMKNE 345  
 Db 1138 PRGSVVTNLYNSTLYEGTKFIKKYAGSNEINVRNDRVYINVVAKNKEYRL----- 1192  
 QY 346 TMDADNHESTDIYAIGUREQTKDINDINIFQI-----QPMNNTYYVASQIFKSNFNG 397  
 Db 1193 ---ATNASQAGVEKILSALEIPDVG--LSQVVMKSKDDQGIKRNK-----CKMNLQD 1240  
 QY 398 ENISGICISGTYRF-----RLGGDYRHHYLVPTVKQNYASLLESTSTHWGFVPVSE 450  
 Db 1241 NNGNDIGFIFGLYDNIKLVASNWN-----RQVGKAS--RTFGCSWEFIPVDD 1288

RESULT 9  
 BXF\_CLOBO STANDARD; PRT; 1274 AA.  
 AC P30996;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BONT/F)  
 DE (Bontoxilysin F).  
 DE BONT.  
 DE Clostridium botulinum.  
 DE Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 DE Clostridium.  
 DE NCBI\_TaxID=1491;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=ATCC 23387;  
 MEDLINE=93012902; PubMed=1398040;  
 East A.K., Richardson P.T., Allaway D., Collins M.D.,

RA Roberts T.A., Thompson D.E.;  
 RT "Sequence of the gene encoding type F neurotoxin of Clostridium  
 RL botulinum";  
 RN FEBS Microbiol. Lett. 75:225-230(1992).  
 RN [2]  
 RC SEQUENCE OF 1-64 FROM N.A.  
 RA STRAIN=Hobbs FT10;  
 RX MEDLINE=94297488; PubMed=7764998;  
 RA East A.K., Collins M.D.;  
 RT "Conserved structure of genes encoding components of botulinum  
 RT neurotoxin complex M and the sequence of the gene coding for the  
 RT nontoxic component in nonproteolytic Clostridium botulinum type F";  
 RL Curr. Microbiol. 29:69-77(1994).  
 RN [3]  
 RP SEQUENCE OF 634-1002 FROM N.A.  
 RX MEDLINE=94013372; PubMed=8408542;  
 RA Campbell K., East A.K., Collins M.D.;  
 RT "Gene probes for identification of the botulinum neurotoxin gene and  
 RT specific identification of neurotoxin types B, E, and F";  
 RL J. Clin. Microbiol. 31:2255-2262(1993).  
 RN [4]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94230352; PubMed=8175689;  
 RA Yamasaki S., Baumeister A., Binz T., Blas J., Link E., Cornille F.,  
 RT Roques B., Fyke E.M., Suedhof T.C., Jahn R., Niemann H.;  
 RN "Cleavage of members of the synaptobrevin/VAMP family by types D and  
 RN F botulinum neurotoxins and tetanus toxin";  
 RL J. Biol. Chem. 269:12764-12772(1994).  
 RN [5]  
 RP FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 RT RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 RT AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 RT WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 RT INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 RT ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-LYS-59  
 RT BOND OF SYNAPTOSOMAL-1 AND -2.  
 RN [6]  
 RP CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 RT neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 RT detected action on small molecule substrates.  
 RN [7]  
 RP SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 RT HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 RT WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 RT FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 RN [8]  
 RP SUBCELLULAR LOCATION: Secreted.  
 RN [9]  
 RP MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 RT BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 RN [10]  
 RP SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 RN [11]  
 RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
 RT between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 RT the European Bioinformatics Institute. There are no restrictions on its  
 RT use by non-profit institutions as long as its content is in no way  
 RT modified and this statement is not removed. Usage by and for commercial  
 RT entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 RT or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 RN [12]  
 RP EMBL; M92906; AAC23263.1; -;  
 DR EMBL; S73676; AAC60475.1; -;  
 DR EMBL; X70820; CAA50151.1; -;  
 DR EMBL; X70816; CAA50147.1; -;  
 DR HSSP; P10845; 3BTA.  
 DR MEPROS; M27.002; -;  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR000130; Zn\_MTPeptide.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR PRODOM; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 FT CHAIN 1 436  
 FT CHAIN 437 1274  
 FT METAL 227 227  
 FT ACT\_SITE 228 228  
 FT METAL 231 231  
 FT ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ZINC (CATALYTIC) (BY SIMILARITY).

FT DISULFID 429 445 INTERCHAIN (PROBABLE).  
 SQ SEQUENCE 1274 AA; 146709 MW; 5B99756A7438B921 CRC64;  
 Query Match 23.08; Score 549.5; DB 1; Length 1274;  
 Best Local Similarity 31.68; Pred. No. 1.7e-25;  
 Matches 147; Conservative 83; Mismatches 180; Indels 55; Gaps 15;  
 QY 2 TIPNIFSYTNNSLLKDIINEYFNNDKSLNSQNKNTPLVDTSGYNAYVSEEGDVQLN 61  
 DB 840 SIPELSSYTNKILIIYFNRLYKIKKIDSLDMRYENKFKIDISGVGNSISNGWYIY 899  
 QY 62 PIFPDKLSSGGEDRGKVIYQENIVYNSMYESISISFWIRNKW---VSNLPGYTII 118  
 DB 900 STNRNQF--GIYNSRLSEVTAQNDIIRYQNFESISISFWIRPKHYKPMNHNRETTII 957  
 QY 119 DSV-KNNSGWSIG---IISNPLVETLQKQEDSQSINFSYDISNNAPGY-NKWFYVTIN 173  
 DB 958 NCMGNNNSGWSIKSLRTVDCHEIIWTLQDTSSNKENLIFRYEELNRIISNINKWIFVTIIN 1017  
 QY 174 NMGMNKIYINGKLIDIKVKELTGINFSTIIFTEINKIPDTGLITSDSDNTINMWIRDEY 233  
 DB 1018 NRLGNSRIYINGNLIVEKSIISNLGDIHVSQNLIFK-----IVGDDETVYVGRIFK 1068  
 QY 234 IFARELQGDKNILFNLSQXTNVVVKDYGNDLRYNKKEYVMVNDIYLRMYANSRQIVFN 293  
 DB 1069 VENTELDKTEIETIYSNEPQPSILKNYGNLYLYNKYILFNLRLKDKYITLNSGILIN 1128  
 QY 234 TRN-----NNDNFEGYKIIIR---TRGNTDTRVGGDILYFDYTNKKAYNLFM 342  
 DB 1129 QQRGTGSEVFLNKLKLYGEVEIIRKNGPIDISNTDNFVAKNDLAYINVVDGVEYRL-- 1186  
 QY 343 KNECMYADNHSTEDIYAGLRQPKDINDNIIQIOPMNNYVYASQIFKSPNGENISG 402  
 DB 1187 -----YADTKSEKELI-----RTSNLDS-LQIIVMDSIGNNCTMNFQNN-NGSN--- 1231  
 QY 403 ICSTGYRFR-----LGGDWYRHNYLVPTVKGQNVASLLESTSHW 443  
 DB 1232 ---IGLLGFHNNLVAWSWYNNIRRTSSNGCFWSSI-SKENGW 1272  
 RESULT 10  
 TETX\_CLOTE STANDARD; PRT; 1314 AA.  
 ID TETX\_CLOTE  
 AC P04958;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Tetanus toxin precursor (EC 3.4.24.68) (tontoxylisin).  
 OS Clostridium tetani.  
 OG Bacterioides.  
 OC Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87053814; PubMed=3536478;  
 RA Eisel U., Jarausch W., Goretzki K., Henschen A., Engels J.,  
 RA Weller U., Rudel M., Habermann E., Niemann H.;  
 RT "Tetanus toxin: primary structure, expression in E. coli, and  
 RT homology with botulinum toxins";  
 RL EMBO J. 5:2495-2502(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CRS911;  
 RX MEDLINE=87040747; PubMed=3774547;  
 RA Fairweather N.F., Lyness V.A.;  
 RT "The complete nucleotide sequence of tetanus toxin";  
 RL Nucleic Acids Res. 14:7809-7812(1986).  
 RN [3]  
 RP SEQUENCE OF 742-1314 FROM N.A.  
 RX MEDLINE=86085672; PubMed=3510187;  
 RA Fairweather N.F., Lyness V.A., Pickard D.J., Allen G., Thomson R.O.;  
 RT "Cloning, nucleotide sequencing, and expression of tetanus toxin

fragment C in *Escherichia coli*.  
 J. Bacteriol. 165:21-27(1988).  
 [4] PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 MEDLINE-90201034; PubMed-2108021;  
 Krieglstein K., Henschen A.H., Weller U., Habermann E.;  
 "Arrangement of disulfide bridges and positions of sulphydryl groups  
 in tetanus toxin."  
 Eur. J. Biochem. 188:39-45(1990).  
 [5] PARTIAL SEQUENCE.  
 MEDLINE-92037649; PubMed-1935979;  
 Krieglstein K.G., Henschen A.H., Weller U., Habermann E.;  
 "Limited proteolysis of tetanus toxin. Relation to activity and  
 identification of cleavage sites."  
 Eur. J. Biochem. 202:41-51(1991).  
 [6] IDENTIFICATION AS ZINC-PROTEASE.  
 MEDLINE-93010948; PubMed-1396558;  
 Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,  
 Montecucco C.;  
 "Tetanus toxin is a zinc protein and its inhibition of  
 neurotransmitter release and protease activity depend on zinc."  
 EMBO J. 11:3577-3583(1992).  
 [7] IDENTIFICATION OF SUBSTRATE.  
 MEDLINE-93063293; PubMed-1331807;  
 Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,  
 Dasgupta B.R., Montecucco C.;  
 "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 by proteolytic cleavage of synaptobrevin."  
 Nature 359:832-835(1992).  
 [8] X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.  
 MEDLINE-97475217; PubMed-9334741;  
 Umland T.C., Winger L.M., Swaminathan S., Furey W.F., Schmidt J.J.,  
 Sax M.;  
 "Structure of the receptor binding fragment HC of tetanus  
 neurotoxin."  
 Nat. Struct. Biol. 4:788-792(1997).  
 [9] FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77  
 BOND OF SYNAPTOSOMAL-2.  
 [10] CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN  
 SYNAPTOSOMAL-2.  
 [11] SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO  
 YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE  
 AND ARE NON-TOXIC AFTER SEPARATION.  
 [12] MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO  
 GANGLIOSIDE RECEPTORS.  
 [13] SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X04436; CAA28033.1; -  
 EMBL; M12739; AAA23282.1; -  
 EMBL; X06214; CAA29564.1; -  
 PIR; A25689; BTCLTN  
 PDB; 1A99; 29-APR-98.  
 PDB; 1A8D; 14-OCT-98.  
 MEROPS; M27.001; -  
 InterPro: IPR000395; Bontoxilysin.  
 InterPro: IPR000130; Zn\_Mtpeptidse.

DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PRO0760; BONTOXILYSIN.  
 DR PRODOM; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;  
 KW 3D-structure.  
 FT INIT\_MET 0  
 FT CHAIN 1 456  
 FT CHAIN 457 1314  
 FT METAL 232 232  
 FT ACT\_SITE 233 233  
 FT METAL 236 236  
 FT TRANSMEM 236 246  
 FT TRANSMEM 669 689  
 FT DISULFID 438 466  
 FT DISULFID 1076 1092  
 SQ SEQUENCE 1314 AA; 150550 MW; 134C3657133EF81D CRC64;  
 Query Match 17.8%; Score 425; DB 1; Length 1314;  
 Best Local Similarity 26.4%; Pred. No. 4.5e-18;  
 Matches 134; Conservative 90; Mismatches 169; Indels 114; Gaps 21;  
 QY 3 IPP-...NIFSTNNSLLKDIINEYFENNINDSKILSLQNRKNTLVDTSGYNAEYSEGDV 58  
 DB 858 IPEYSKNDLCWVDNEEDIDVI-----LKRSTILNIDINDISDIGNSSVITYPDA 911  
 QY 59 QLNPIEPFDFKLGSGE-----DRGKIVIVTQENIYVNSMYESFSISFWIRNK-WVS 110  
 DB 912 QLPV-----GINKAHLVNNESSEVIVHKAMDIEYNDMNFNTVSWLRVPKVSAS 963  
 QY 111 NLP-...GYTIDSVKNN-----SGWSIGIISNLFVTLKQNEDESGSINFSYDISNNA 160  
 DB 964 HLEOYCTNEYSIISNMKHSLSIGSGWSVSLKGNLITLKDASAGEVQITFR-DLPDKF 1022  
 QY 161 PGV--NKWFEVTVTNMGMNMIYKGLIDITIKVKEITGINSFKTIIFTEINKIPDTGLI 218  
 DB 1023 NAYLANKWFIITNDRLSSANLYNGVLMSAEITGLGAIREDNNTIKLDR----- 1075  
 QY 219 TSQSDNNKWRDVFIFAKELGDKDINILFNSLQTVNVVQYWGNDLYRKEYYMWNI-- 276  
 DB 1076 -CNNOYYSIDKFRICFALPKPEIKLYTSITFLRDFWGNPLRYDTEYLLIPVAS 1134  
 QY 277 -----DYL---NRYMANSRQIVNFRNRNNDNFEGYKIIKIRGNTN-DTRV 321  
 DB 1135 SSKDVOLKNITDYMLTNAPSYTNGKLIYYRLYN-----GLFKIKRYTPNEIDSV 1189  
 QY 322 RGGDILYFMTINNAY-----NLFMKNMTYADNHSTEDIYAGLREQTKDINDNI 373  
 DB 1190 KSGDFIKLYSVYNNNEHIVGYPKDGNFANLDRILRVGYNAPGIPLYKKMEAVK----- 1243  
 QY 374 IFQIQPMNNTYYASQIFKSNFNGENISGICSGYRFRLGSD-----WYRHNLYL 423  
 DB 1244 -----LRDLKTYSVQL--KLYDDKNAS-LGLVGHNGIGNDPRDLIASNRY-FNHL 1293  
 QY 424 VPTVKOGNYASLLESTSTHWGFVPVSE 450  
 DB 1294 KDXI-----LGCDWYFVPTDE 1309  
 RESULT 11  
 ID EXEN\_CLOBO STANDARD; PRT; 1162 AA.  
 AC P46082;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Botulinum neurotoxin type E, nontoxic component.  
 GN BNT-120  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=Maslike;
RA MEDLINE=93195515; PubMed=8450310;
RA Fujii N., Kimura K., Yokosawa N., Yashiki T., Tezuka K., Oguma K.;
RT "The complete nucleotide sequence of the gene encoding the nontoxic
RT component of Clostridium botulinum type E progenitor toxin.";
RL J. Gen. Microbiol. 139:79-86(1993).
CC
CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D12697; BAA02194.1;
CC InterPro: IPR000395; Bontotoxilisn.
CC Pfam: PF01742; Peptidase_M27; 1.
CC PRINTS; PR00760; BONTOTOXILYSIN.
CC ProDom; PD001963; Bontotoxilisn; 1.
CC Neurotoxin.
CC
CC SEQUENCE 1162 AA; 136856 MW; 96468EDDAAE0F39D CRC64;
CC
CC Query Match 9.68; Score 230.5; DB 1; Length 1162;
CC Best Local Similarity 22.88; Pred. No. 1.5e-06;
CC Matches 96; Conservative 74; Mismatches 158; Indels 93; Gaps 17;
CC
CC 6 NIFSTNNSLLKDIINEYFNNDKLSLQNRKNTLV-DTSGYNAEYSEGDVQLNP1F 64
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 783 NLFNLYTELLIKEQTSFY-----ELSLYAFQEDNNVIGDTSGKNTLVE-----Y 827
CC
CC 65 PFDKLGSSGDRGKVIQTQENIVY-----NSMYSPSISFWIRINKWVSNLPGYTI- 117
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 828 PKDIGLVYGINNAHLTGANQKFNDFENGLTNFSLYFWLR-----NLKONTIK 881
CC
CC 118 ---IDSVKNSGWSIGIISNPLVFTLQKQEDSEOSINFSYDISNAPGYNKFFVTN 173
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 882 SKLIGSKEDNCGWEIFYFENGLVFNIDSGNEKNIVLS-NISNKS-----WHYIVISIN 935
CC
CC 174 NMGMNKIYINGKLDTTKVKELGTGINSKTIPEINKIPDTGLTSDSNINWIRDFY 233
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 936 RLKQDLIFIDNLVANEDKEILNIYSS-----DIISLSDNNNVYIEGLS 982
CC
CC 234 IFAPKELGDKDINILFNSLQYTNVVKDWGNDLRYNKYYVMYNDY----LNRYMYANSRQ 289
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 983 VLNKTINSNEILDTYFSLDNNYSIRNDEEILQYNRTYELFNVPFPIAINKIEQNNIY 1042
CC
CC 290 IVENTRRNNDFNKKIIRKIRGNTNDR-----VRGGDILYFDWNTINKA 337
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 1043 LSNINNENLNFKPLKFKLL-----NTNPNKQYQKWDVEIFSVLDGTEKYLDISTN-- 1094
CC
CC 338 YNLFMKNETMYADNHSTEDIIAIGLRQETKDINDNIIFQOPM--NNTYYIASQIFKSNF 395
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 1095 -----NRIQLVDNKNNAKFTIIN-----NDIFISNCLTLTYNNVNVYLS- 1139
CC
CC 396 N 396
CC
CC 1140 N 1140
CC
CC RESULT 12
CC BYEN_CLOBU STANDARD; PRT; 1162 AA.
CC ID EXCN_CLOBU
CC AC Q06366;
CC DT 01-FEB-1995 (Rel. 31, Created)
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
CC DT 01-FEB-1995 (Rel. 32, Last annotation update)
CC DE Botulinum neurotoxin type E, nontoxic component.
CC OS Clostridium botulinum.
CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

```

```

CC Clostridium.
CC NCBI_TaxID=1492;
CC (1)
CC RP SEQUENCE FROM N.A.
CC STRAIN=BL6340;
CC MEDLINE=93360835; PubMed=8355622;
CC RA Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
CC Takeshi K., Ohya T., Isogai E., Isogai H.;
CC "Similarity in nucleotide sequence of the gene encoding nontoxic
CC component of botulinum toxin produced by toxigenic Clostridium
CC botulinum strain BL6340 and Clostridium botulinum type E strain
CC Maslike.";
CC Microbiol. Immunol. 37:395-398(1993).
CC
CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D12739; BAA02231.1;
CC InterPro: IPR000395; Bontotoxilisn.
CC Pfam: PF01742; Peptidase_M27; 1.
CC PRINTS; PR00760; BONTOTOXILYSIN.
CC ProDom; PD001963; Bontotoxilisn; 1.
CC Neurotoxin.
CC
CC SEQUENCE 1162 AA; 136829 MW; C86E9BE932DA78E4 CRC64;
CC
CC Query Match 9.68; Score 230.5; DB 1; Length 1162;
CC Best Local Similarity 22.98; Pred. No. 1.5e-06;
CC Matches 97; Conservative 76; Mismatches 152; Indels 99; Gaps 19;
CC
CC 6 NIFSTNNSLLKDIINEYFNNDKLSLQNRKNTLV-DTSGYNAEYSEGDVQLNP1F 64
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 783 NLFNLYTELLIKEQTSFY-----ELSLYAFQEDNNVIGDTSGKNTLVE-----Y 827
CC
CC 65 PFDKLGSSGDRGKVIQTQENIVY-----NSMYSPSISFWIRINKWVSNLPGYTI- 117
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 828 PKDIGLVYGINNAHLTGANQKFNDFENGLTNFSLYFWLR-----NLNQTNIK 881
CC
CC 118 ---IDSVKNSGWSIGIISNPLVFTLQKQEDSEOSINFSYDISNAPGYNKFFVTN 173
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 882 SKLIGSKEDNCGWEIFYFENGLVFNIDSGNEKNIVLS-NISNKS-----WHYIVISIN 935
CC
CC 174 NMGMNKIYINGKLDTTKVKELGTGINSKTIPEINKIPDTGLTSDSNINWIRDFY 233
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 936 RLKQDLIFIDNLVANEDKEILNIYSS-----DIISLSDNNNVYIEGLS 982
CC
CC 234 IFAPKELGDKDINILFNSLQYTNVVKDWGNDLRYNKYYVMYNDY----LNRYMYANSRQ 289
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 983 VLNKTINSNEILDTYFSLDNNYSIRNDEEILQYNRTYELFNVPFPIAINKIEQNNN-- 1040
CC
CC 290 IVENTRRNNDFN---EYKIIIRKIRGNTNDR-----VRGGDILYFDWNTIN 334
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 1041 -IYLSNNNENLNFKPLKFKLL-----NTNPNKQYQKWDVEIFSVLDGTEKYLDISID 1093
CC
CC 335 NKAYNLFMKNETMYADNHSTEDIIAIGLRQETKDINDNIIFQOPM--NNTYYIASQIFK 392
CC :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
CC 1094 N-----NRIQLVDNKNNAKFTIIN-----NDIFISNCLTLTYNNVNVYLS- 1136
CC
CC 393 SNFN 396
CC :||:
CC 1137 QDYN 1140
CC
CC RESULT 13
CC BYCN_CLOBU
CC ID EXCN_CLOBO STANDARD; PRT; 1196 AA.
CC AC P46081;

```



Db 382 -NKNGK 386

RESULT 15

ID C7AB\_BACUK STANDARD; PRT: 1138 AA.  
AC Q45708;  
DT 16-CCN-2001 (Rel. 40, Created)  
DT 16-CCN-2001 (Rel. 40, Last sequence update)  
DT 16-CCN-2001 (Rel. 40, Last annotation update)  
DE Pesticidal crystal protein cry7AB (insecticidal delta-endotoxin  
DE CryIIA(b)) (crystalline entomocidal protoxin) (130 kDa crystal  
DE protein).  
GN CRY7AB OR CRYVIIA(B).  
OS Bacillus thuringiensis (subsp. Kumamotoensis).  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

NCBI\_TaxID-132267;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=HD867;  
Payne J.M., Fu J.M.;  
"Coleopteran-active Bacillus thuringiensis isolates and genes encoding  
coleopteran-active toxins.";  
Patent number US5286486, 15-FEB-1994.  
-!- FUNCTION: PROMOTES COLLOIDSMORIC LYSIS BY BINDING TO THE MIDGUT  
EPITHELIAL CELLS OF COLEOPTERA.  
-!- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING  
SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART  
OF THE SPORE COAT.  
-!- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE  
N-TERMINUS.  
-!- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; U04368; AAA21121.1; .  
HSSP; P07130; 1DLC.  
InterPro; IPR001178; Endotoxin.  
Pfam; PF00555; endotoxin; 1.  
Toxin; Sporulation.  
SEQUENCE 1138 AA; 129658 MW; E12DC80CA56D1DA CRC64;

Query Match 6.5%; Score 155; DB 1; Length 1138;  
Best Local Similarity 23.1%; Pred. No. 0.044;  
Matches 109; Conservative 56; Mismatches 134; Indels 172; Gaps 25;

2 TTPFNIRSYTNLSLKDIINEYFNINDSKILSLQNKNTLVDTSGYNAEVSEGDVQLN 61  
475 ITP--IFSWTHSA-----EYNIPTNKTIPAVKM----- 505  
62 PTFPDKLSSGEDRGKVIIVTQENIVNVMYSEFSISFWIRINKWVNLPGYTIIDSV 121  
506 -----YKLGDTST-----VYKGPFTGGDLV 526  
122 KNSGWSIGISNFIIVTLKONDESEOSINFSDISNNAPGNKWFVTVNNMGMNKI 181  
527 KRGSNGYIGDKAVNPSLQN-----YAVR-----VRYATNVSGQFNW 565  
182 YINGKL-----IDTI-KVKELT-----GINFSKTIITFEINKIPDTGLITSDSDNIN 226  
566 YINDKITLQKQFQNTVEIGBGKLTGSGFYIEYSTIQFP-DKHPKITLHLSDLNNS 624  
227 MWIRDFYIFAKEL-----DGKD-----INILF-----YTNVVK 258  
625 ----SFYDYSIEFIPDVNVDYDEKLEKAKAVNTLEGRNALQKDYTDYKVDQVSILV 680

QY 259 DYWGNDLRNKKEYTMVN-IDYLNRYMYANSQIVFN---TRNNNDNFEGY---KIIIKR 311  
Db 681 DCISGDLYPNEKRELQNLVYAKRLSY--SRNLLDPTFDSINSSENGWYGSNGIVGN 738  
QY 312 ----IRGN-----TNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADN---HSTEDIY 358  
Db 739 GDFVFKGNLYLIFSGTNDTQ-----YPTLYQKIDESKLEKEYTRYKLGFTESSQDLE 790  
QY 359 AIGLR---BOTKDINDNIIFQIQPMNTVYVYASQIFKSNFNENISGICS 405  
Db 791 AYVIRYDAKHRTLDVSNLLPDILP-ENTCGEPNRCAAQOYLDENPSSECS 840

Search completed: November 7, 2002, 14:44:02  
Job time : 34 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:40:06 ; Search time 81 Seconds  
(without alignments)  
1144.706 Million cell updates/sec

Title: US-09-910-186A-10  
Perfect score: 2391  
Sequence: 1 TTFNFSTNNLLKDI.....NYASLLESTHWGFPVSE 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp-organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp Vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2386	99.8	1291	2 Q93HT3	Q93HT3 clostridium
2	1824	76.3	1285	2 Q9LBR1	Q9LBR1 clostridium
3	1823	76.2	1285	2 Q45967	Q45967 clostridium
4	882.5	36.9	1280	2 Q9LBS7	Q9LBS7 clostridium
5	882.5	36.9	1280	2 Q45849	Q45849 clostridium
6	867.5	36.3	1275	12 Q9QTC7	Q9QTC7 clostridium
7	610	25.5	1291	2 Q933K0	Q933K0 clostridium
8	604	25.3	1291	2 Q9ZAU8	Q9ZAU8 clostridium
9	602.5	25.2	1255	2 Q9FAR6	Q9FAR6 clostridium
10	602	24.8	1291	2 Q93G71	Q93G71 clostridium
11	594	24.8	1251	2 Q9K395	Q9K395 clostridium
12	593	24.8	1291	2 Q9Q077	Q9Q077 clostridium
13	575	24.0	441	2 Q9X708	Q9X708 clostridium
14	569	23.8	1268	2 Q45851	Q45851 clostridium
15	564	23.6	1278	2 Q57236	Q57236 clostridium
16	527.5	22.1	1280	2 Q9ZAU5	Q9ZAU5 clostridium

17	428	17.9	451	2	Q9LAI3	Q9LAI3 clostridium
18	428	17.9	1310	2	Q93N27	Q93N27 clostridium
19	428	17.9	361	2	Q45848	Q45848 clostridium
20	312	13.0	361	2	Q45846	Q45846 clostridium
21	281.5	11.8	367	2	Q45861	Q45861 clostridium
22	281.5	11.8	367	2	Q45862	Q45862 clostridium
23	258	10.8	1197	2	Q45888	Q45888 clostridium
24	239.5	10.0	1198	2	Q06018	Q06018 clostridium
25	236.5	9.9	1197	2	Q33871	Q33871 clostridium
26	236.5	9.9	1197	2	Q9ZAJ9	Q9ZAJ9 clostridium
27	234	9.8	1197	2	Q9ZAJ9	Q9ZAJ9 clostridium
28	231.5	9.7	1198	2	Q69277	Q69277 clostridium
29	230.5	9.6	1198	2	Q45893	Q45893 clostridium
30	216.5	9.1	1193	2	Q45880	Q45880 clostridium
31	215.5	9.0	1161	2	Q45891	Q45891 clostridium
32	215.5	9.0	1161	2	Q69276	Q69276 clostridium
33	213	8.9	1193	2	Q45914	Q45914 clostridium
34	213	8.9	1193	2	Q71107	Q71107 clostridium
35	208	8.7	1162	2	Q9ZAJ6	Q9ZAJ6 clostridium
36	200	8.4	1196	2	Q9LBS8	Q9LBS8 clostridium
37	199	8.3	1165	2	Q45844	Q45844 clostridium
38	199	8.3	1165	2	Q45887	Q45887 clostridium
39	199	8.3	1196	2	Q45916	Q45916 clostridium
40	199	8.3	1196	2	Q93HT4	Q93HT4 clostridium
41	199	8.3	1196	9	Q38197	Q38197 clostridium
42	197	8.2	1163	2	Q45850	Q45850 clostridium
43	197	8.2	1196	2	Q9LBR2	Q9LBR2 clostridium
44	197	8.2	1196	2	Q53550	Q53550 clostridium
45	197	8.2	1196	9	Q9ZXX7	Q9ZXX7 clostridium

## ALIGNMENTS

RESULT 1					
Q93HT3					
ID	Q93HT3	PRELIMINARY;	PRT;	1291 AA.	
AC	Q93HT3;				
DT	01-DEC-2001 (TRENBLrel. 19, Created)				
DT	01-DEC-2001 (TRENBLrel. 19, Last sequence update)				
DT	01-MAR-2002 (TRENBLrel. 20, Last annotation update)				
DE	Neurotoxin.				
GN	Nr.				
OS	Clostridium botulinum.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;				
OC	Clostridiales; Clostridiaceae; Clostridium.				
OX	NCBI_TaxID=1491;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-C-YOICHI;				
RA	MEDLINE-21534265; PubMed=11676492;				
RA	Sagane Y., Kouguchi H., Watanabe T., Sunagawa H., Inoue K.,				
RA	Fujinaga Y., Oguma K., Ohyama T.;				
RT	"Role of C-Terminal Region of HA-33 Component of Botulinum Toxin in				
RT	Hemagglutination."				
RL	Biochem. Biophys. Res. Commun. 288:650-657(2001).				
DR	EMBL; AB061780; BAB1749.1; .				
DR	InterPro; IPR000395; Bontoxilysin.				
DR	InterPro; IPR000130; Zn_MTPeptide.				
DR	Pfam; PF01742; Peptidase_M27; 1.				
DR	ProDom; PD001963; Bontoxilysin; 1.				
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN1.				
DR	SEQUENCE 1291 AA; 148869 MW; 4A21DB35B8743CF8 CRC64;				

Query Match 99.8%; Score 2386; DB 2; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 5.7e-117;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTFNFSTNNLLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEVSSEGDVQLN 61

Db 843 TTFNFSTNNLLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEVSSEGDVQLN 902

QY 62 PTFPFKLGSSGDRGVIVTQENIYVNSYFSISFWIRNKWVNLPGYTIIDS 121

```
Db 903 PIFPFDKLGSSGDRGKVIYQENIVYNSMYESFSISFWIRKWSNLPGYTIIDSV 962
Qy 122 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDINSNAPGYNKWFVTVTNMGNMKI 181
Db 963 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDINSNAPGYNKWFVTVTNMGNMKI 1022
Qy 182 YINGKLIDTIVKVELTGINFSEKTIITFEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 241
Db 1023 YINGKLIDTIVKVELTGINFSEKTIITFEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 1082
Qy 242 KDINILFNSLOYTVNVKDYWGNDLRYNKEYVMYVNIIDYLRNRYMYANSRQIVFTRNNNDP 301
Db 1083 KDINILFNSLOYTVNVKDYWGNDLRYNKEYVMYVNIIDYLRNRYMYANSRQIVFTRNNNDP 1142
Qy 302 NEGKIIIRKIRGTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSHTEDIYATG 361
Db 1143 NEGKIIIRKIRGTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSHTEDIYATG 1202
Qy 362 LREOTKDINDNIIFQIOPMNTYYASQIFKSNFNGENISGICSTGYRFRLLGGDWYRHN 421
Db 1203 LREOTKDINDNIIFQIOPMNTYYASQIFKSNFNGENISGICSTGYRFRLLGGDWYRHN 1262
Qy 422 YLVPTVKQGNVYASLLESTSTHWGFPVSE 450
Db 1263 YLVPTVKQGNVYASLLESTSTHWGFPVSE 1291

RESULT 2
ID Q9LBR1 PRELIMINARY; PRT; 1285 AA.
01-OCT-2000 (Tremblrel. 15, Created)
01-OCT-2000 (Tremblrel. 15, Last sequence update)
01-DEC-2001 (Tremblrel. 19, Last annotation update)
Neurotoxin.
NT.
Clostridium botulinum.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
Clostridiales; Clostridiaceae; Clostridium.
NCBI_TaxID=1491;
[1]
SEQUENCE FROM N.A.
STRAIN=D-4947;
Sagane Y., Watanabe T., Kouguchi H., Yamamoto T., Takizawa J.,
Kawabe T., Murakami F., Muroga A., Nakatsuka M., Ohyama T.;
"Characterization of the progenitor toxin components produced by
Clostridium botulinum Type D Strain 4947";
Submitted (28B-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AB037920; BAA90661.1; -
HSP; P10845; 3BTA.
MEROPS; M27.002; -.
InterPro; IPR000395; Bontoxilysin.
Pfam; PF01742; Peptidase_M27; 1.
PRINTS; PR00760; BONTOXILYSIN.
ProDom; PD001963; Bontoxilysin; 1.
PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1285 AA; 147352 MW; B63AFA487D570680 CRC64;

Query Match 76.3%; Score 1824; DB 2; Length 1285;
Best Local Similarity 77.9%; Pred. No. 1.3e-87;
Matches 352; Conservative 36; Mismatches 56; Indels 8; Gaps 4;

Qy 2 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61
Db 839 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 898
Qy 62 PIFPFDKLGSSGDRGKVIYQENIVYNSMYESFSISFWIRKWSNLPGYTIIDSV 121
Db 899 PIFPFDKLGSSGDRGKVIYQENIVYNSMYESFSISFWIRKWSNLPGYTIIDSV 958
Qy 122 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDINSNAPGYNKWFVTVTNMGNMKI 181
Db 963 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDINSNAPGYNKWFVTVTNMGNMKI 1022
```

```
Db 959 KNSGWSIGIISNLFVTLKQNESEQDINFSYDISKNAAGYNKWFVTVTNMGNMKI 1018
Qy 182 YINGKLIDTIVKVELTGINFSEKTIITFEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 241
Db 1019 YINGKLIDTIVKVELTGINFSEKTIITFEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 1078
Qy 242 KDINILFNSLOYTVNVKDYWGNDLRYNKEYVMYVNIIDYLRNRYMYANSRQIVFTRNNNDP 301
Db 1079 KDINILFNSLOYTVNVKDYWGNDLRYNKEYVMYVNIIDYLRNRYMYANSRQIVFTRNNNDP 1138
Qy 302 NEGKIIIRKIRGTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSHT-EDIYAT 360
Db 1139 NEGKIIIRKIRGTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSHT-EDIYAT 1193
Qy 361 GLREOTKD-INDNIIFQIOPMNTYYASQIFKSNFNGENISGICSTGYRFRLLGGD-WY 418
Db 1194 GALDQPMDETRKYGSTFIQPCNTFDYASQLFLSNATNRLGILSIGSYSPKLGDDYWF 1253
Qy 419 RHNLYVTVKQGNVYASLLESTSTHWGFPVSE 450
Db 1254 NHEYLPIVIEHYASLLESTSTHWGFPVSE 1285

RESULT 3
ID Q45967 PRELIMINARY; PRT; 1285 AA.
AC Q45967;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Neurotoxin CONSISTING of botulinum neurotoxin D and Ci.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
NCBI_TaxID=1491;
[1]
SEQUENCE FROM N.A.
STRAIN-TYPE D;
MEDLINE=96283801; PubMed=8679691;
Moriishi K., Koura M., Abe N., Fujii N., Fujinaga Y., Inoue K.,
Ogumada K.;
"Mesal structures of neurotoxins produced from Clostridium botulinum
Types C and D organisms";
Biochim. Biophys. Acta 1307:123-126(1996).
EMBL; D38442; BAA07477.1; -.
HSP; P10845; 3BTA.
MEROPS; M27.002; -.
InterPro; IPR000395; Bontoxilysin.
Pfam; PF01742; Peptidase_M27; 1.
PRINTS; PR00760; BONTOXILYSIN.
ProDom; PD001963; Bontoxilysin; 1.
PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Neurotoxin.
SQ SEQUENCE 1285 AA; 147365 MW; EDE98E4E6A6C413 CRC64;

Query Match 76.2%; Score 1823; DB 2; Length 1285;
Best Local Similarity 77.7%; Pred. No. 1.5e-87;
Matches 351; Conservative 37; Mismatches 56; Indels 8; Gaps 4;

Qy 2 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61
Db 839 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 898
Qy 62 PIFPFDKLGSSGDRGKVIYQENIVYNSMYESFSISFWIRKWSNLPGYTIIDSV 121
Db 899 PIFPFDKLGSSGDRGKVIYQENIVYNSMYESFSISFWIRKWSNLPGYTIIDSV 958
Qy 122 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDINSNAPGYNKWFVTVTNMGNMKI 181
Db 959 KNSGWSIGIISNLFVTLKQNESEQDINFSYDISKNAAGYNKWFVTVTNMGNMKI 1018
```



QY 182 YINGKLIDITIKVKGELTGINFSTIIFKINPKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241  
 Db 1019 YINGKLIDITIKVKGELTGINFSTIIFQNKIPNTGLITSDSDNINMWIRDFYIFAKELDD 1078  
 QY 242 KINILFNSLQTNVVKYWGNDLRVKNKEYYVWIDYLNRYMYANSROIVFTRNNN 301  
 Db 1079 KINILFNSLQTNVVKYWGNDLRVKNKEYYVWIDYLNRYMYANSROIVFTRNNN 1138  
 QY 302 NEGYKIIIRKIRGNTNDTRVRGDDILYFDMTNNKAYNLFMKNETMYADNHST-EDIYAI 360  
 Db 1139 NEGYKIIIRKIRGNTNDTRVRGDDILYFDMTNNKAYNLFMKNETMYADNHST-EDIYAI 1193  
 QY 361 GUREQTKD-INDNIIQICPMNNIYVYASQIFKSNFNGENISGICSTGTYFRIGSD-WY 418  
 Db 1194 GUREQTKD-INDNIIQICPMNNIYVYASQIFKSNFNGENISGICSTGTYFRIGSD-WY 1253  
 419 RENVLPVTKQGNVYASLLESTSTHGWFPVSE 450  
 1254 NREYLPVTKQGNVYASLLESTSTHGWFPVSE 1285  
 RESULT 4  
 Q9LBS7 PRELIMINARY; PRT; 1280 AA.  
 01-OCT-2000 (TREMELrel. 15, Created)  
 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
 Neurotoxin.  
 NT.  
 Clostridium botulinum.  
 Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 Clostridiales; Clostridiaceae; Clostridium.  
 NCBI\_TaxID=1491;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=C-6814;  
 Sagane Y., Watanabe T., Kouguichi H., Yamamoto T., Kawabe T.,  
 Murakami F., Nakatsuka M., Ohyama T.;  
 "Organization of Gene Encoding Components of the Botulinum Progenitor  
 Toxin in Clostridium botulinum Type C Strain 6814: Evidence of  
 ChimERIC Sequence in the Gene Encoding Each Component.",  
 Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 EMBL; AB037166; BAA69713.1;  
 HSP; P10845; 3BTA.  
 MEROPS; M27.002;  
 InterPro; IPR000395; Bontoxilysin.  
 InterPro; IPR000130; Zn\_Mtpeptidse.  
 Pfam; PF01742; Peptidase\_M27; 1.  
 PRINTS; PR00760; BONTOXILYSIN.  
 ProDom; PD001963; Bontoxilysin; 1.  
 PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 SEQUENCE 1280 AA; 147757 MW; DE124FFB6F68450B CRC64;  
 Query Match 36.9%; Score 882.5; DB 2; Length 1280;  
 Best Local Similarity 43.9%; Pred. No. 2.le-38;  
 Matches 198; Conservative 78; Mismatches 148; Indels 27; Gaps 10;  
 QY 2 TIPFNIFSYNNLSLKDIIINEYFNINDSKILSLQNRKNTLVDTSGYNAEYSEGVDQLN 61  
 Db 843 TIPFNIFSYNNLSLKDIIINEYFNINDSKILSLQNRKNTLVDTSGYNAEYSEGVDQLN 902  
 QY 62 PTFPDKLGSGEDRGKVIYTONENIYVNSMYSFISFWIRNKVSNLPG-YTIIDS 120  
 Db 903 TIYTNDFKLSSGSD---KIIVNLLNIIYSAIYENSSVFWIKSKDLTNSHREYTIINS 959  
 QY 121 VKNNSGWSIGISNFIPLVFTLKONEDSEOSINFSDISNAPGY-NKWFVYVTVNNMGMN 179  
 Db 960 IKQNSGWLKICIRNGIEWLQDINRKYSLFDYSESLSHPTGYNKWFVYVTVNNMGMN 1019  
 QY 180 KYINGKLIDITIKVKGELTGINFSTIIFKINPKIPDTGLITSDSDNINMWIRDFYIFAKEL 239  
 Db 1020 KYINGKLIDITIKVKGELTGINFSTIIFKINPKIPDTGLITSDSDNINMWIRDFYIFAKEL 1071

QY 240 DGKDNILFNSLQTNVVKYWGNDLRVKNKEYYVWIDYLNRYMYANSROIVFTRNNN 299  
 Db 1072 SNEDINIVFEGOILRNKIRKIDYGNPLKFDTEYIINNDYIDRYAPAKSNILVLVOYFDRS 1131  
 QY 300 DFNEGYKIIIRKIRGNTNDTRVRGDDILYFDMTNNKAYNLFMKNETMYA---DNHSTED 356  
 Db 1132 KLYTGNPITIKSVSDKNPYSELNLMFHLNMGKMIIRDTDIYATGREGCSKNC 1191  
 QY 357 IYAIQGLRQETKDINDNIIQICPMNNIYVYASQIFKSNFNGENISGICST-GTYFRIGLG 415  
 Db 1192 IYAIQGLRQETKDINDNIIQICPMNNIYVYASQIFKSNFNGENISGICST-GTYFRIGLG 1246  
 QY 416 DWYRHNLYPVTYKQGNVYASLLESTSTHGWFPV 446  
 Db 1247 DWYRHNLYPVTYKQGNVYASLLESTSTHGWFPV 1272  
 RESULT 5  
 Q45849 PRELIMINARY; PRT; 1280 AA.  
 AC Q45849;  
 DT 01-NOV-1996 (TREMELrel. 01, Created)  
 DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
 DE Neurotoxin.  
 OS Clostridium botulinum C.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 OX NCBI\_TaxID=36828;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=6813;  
 RX MEDLINE=96156810; PubMed=8593068;  
 RA Moriishi K., Koura M., Fujii N., Fujinaga Y., Inoue K., Syuto B.,  
 Oguma K.;  
 "Molecular cloning of the gene encoding the mosaic neurotoxin,  
 composed of parts of botulinum neurotoxin types C1 and D, and PCR  
 detection of this gene from Clostridium botulinum type C organisms.",  
 Appl. Environ. Microbiol. 62:662-667(1996).  
 DR EMBL; D49440; BAA08418.1;  
 DR HSP; P10845; 3BTA.  
 DR MEROPS; M27.002;  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR000130; Zn\_Mtpeptidse.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
 SQ SEQUENCE 1280 AA; 147815 MW; 74F813E228B8C989 CRC64;  
 Query Match 36.9%; Score 882.5; DB 2; Length 1280;  
 Best Local Similarity 43.9%; Pred. No. 2.le-38;  
 Matches 198; Conservative 78; Mismatches 148; Indels 27; Gaps 10;  
 QY 2 TIPFNIFSYNNLSLKDIIINEYFNINDSKILSLQNRKNTLVDTSGYNAEYSEGVDQLN 61  
 Db 843 TIPFNIFSYNNLSLKDIIINEYFNINDSKILSLQNRKNTLVDTSGYNAEYSEGVDQLN 902  
 QY 62 PTFPDKLGSGEDRGKVIYTONENIYVNSMYSFISFWIRNKVSNLPG-YTIIDS 120  
 Db 903 TIYTNDFKLSSGSD---KIIVNLLNIIYSAIYENSSVFWIKSKDLTNSHREYTIINS 959  
 QY 121 VKNNSGWSIGISNFIPLVFTLKONEDSEOSINFSDISNAPGY-NKWFVYVTVNNMGMN 179  
 Db 960 IKQNSGWLKICIRNGIEWLQDINRKYSLFDYSESLSHPTGYNKWFVYVTVNNMGMN 1019  
 QY 180 KYINGKLIDITIKVKGELTGINFSTIIFKINPKIPDTGLITSDSDNINMWIRDFYIFAKEL 239  
 Db 1020 KYINGKLIDITIKVKGELTGINFSTIIFKINPKIPDTGLITSDSDNINMWIRDFYIFAKEL 1071  
 QY 240 DGKDNILFNSLQTNVVKYWGNDLRVKNKEYYVWIDYLNRYMYANSROIVFTRNNN 299

```
Db 1072 SNEDINIVYEQILRNWIKDYWGPNLKFDETEYXIINDYIDRYIAPKSNILVLVQYDPS 1131
Qy 300 DFNEGYKIIIRKRGNTNDVRGSDILYFDMTINKKAYNLFMKNETMYA---DNHSTED 356
Db 1132 KLY:GNPTITKSDKPNYSRLNGDNIMPHLYNSKYMIRDTDTIYALGREGCSKNC 1191
Qy 357 IYATGLRQTKDINDNIIFOQPMNNYYIASOIFKSNFNGENISGICSI-GTYFRPLGG 415
Db 1192 VYALKQSLNGLNYGIG-IFSITKNTVSOKYCSQIFSSFM--KNTMLLADIYKPRWFSF---1246
Qy 416 DWYRHNYLVPTVKQGNVYASLLESTSTHWGFV 446
Db 1247 ---ENAYTP-VATNVTETKLLSTSSFWKFI 1272

RESULT 6
Q90TG7 PRELIMINARY; PRT; 1275 AA.
Q90TG7:
01-MAY-2000 (Tremblrel. 13, Created)
01-MAY-2000 (Tremblrel. 13, Last sequence update)
01-DEC-2001 (Tremblrel. 19, Last annotation update)
NTX (fragment).
NCBI_TaxID=29342;
Clostridium botulinum D bacteriophage.
Viruses.
SEQUENCE FROM N.A.
[1]
STRAIN=1873;
MEDLINE=99017546; PubMed=9802560;
Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa K., Takeshi K.,
Ohyama T., Watanabe T., Inoue K., Oguma K.;
"Molecular composition of the 16S toxin produced by a Clostridium
botulinum type D strain, 1873." (1998).
Microbiol. Immunol. 42:599-605 (1998).
EMBL; AB012112; BAA75084.1; -
HSSP; P10845; 3ETA.
MEROPS; M27.002; -.
InterPro; IPR000395; Bontoxilysin.
InterPro; IPR000130; Zn_MTPeptidase.
Pfam; PF01742; Peptidase_M27; 1.
PRINTS; PR00760; BONTOKILYSIN.
ProDom; PD001963; Bontoxilysin; 1.
PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
NON_TER 1275 1275
SEQUENCE 1275 AA; 146742 MW; 3C50F46C8233E2D6 CRC64;

Query Match 36.3%; Score 867.5; DB 12; Length 1275;
Best Local Similarity 43.2%; Pred. NO. 1.3e-37;
Matches 195; Conservative 83; Mismatches 146; Indels 27; Gaps 11;

2 TIPNIFSYTNSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGYNAEVSDEGDVQLN 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
839 TMPNIFSYTNSLLKDIINEYFNINDSKILSLQNRKNTLVDTSGYNAEVRVGDVQLN 898

Qy 62 PIFPDKLSSGDRGKVIYTONENIVNSMYESFSISFWIRNKW---VSNL--PGY 115
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 899 TIYNDKFLSSGSD--KILVNLNNILYSAIYENSVSFWIKSKDLTNSHNEYIINS 955
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 121 VKNNSGWSIGISNLFVTLTKQNEDEQSINFSYDINSNAPGY-NKWFFVTNNMGNM 179
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 956 IEQSGNKKLCIRNGNLEWLOVDNRKYSKLFYDSELSHTGYTNKWFVTNNMGYK 1015
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 180 KIYNGKLIDIKVKELTGNTFSITFEINKIPDGLTSDSDNNIMWIRYFAKEL 239
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1016 KLYINGLKQSKTDEDEVKLDITVFGDENID-----ENQMLWIRDFNIFSKEL 1067
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 240 DGKDNILFNSLOQTVNVKDYWGNDLRYNKEYYMNIDYLNRYMANSRQIVFTRNNN 299
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1066 SNEDINIVYEQILRNWIKDYWGPNLKFDETEYXIINDYIDRYIAPKSNILVLVQYDPS 1127
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 300 DFNEGYKIIIRKRGNTNDVRGSDILYFDMTINKKAYNLFMKNETMYADNH---STED 356
```

```
Db 1128 KLYTGNDITTKSVDRKPNYSRLNGDNIMHLYNSKYMIRDTDTIYATQGCSONC 1187
Qy 357 IYATGLRQTKDINDNIIFOQPMNNYYIASOIFKSNFNGENISGICSI-GTYFRPLGG 415
Db 1188 VYALKQSLNGLNYGIG-IFSITKNTVSOKYCSQIF-SSFR-ENTMLLADIYKPRFS---1241
Qy 416 DWYRHNYLVPTVKQGNVYASLLESTSTHWGFV 446
Db 1242 --PKNAVTPAVT--NYETKLLSTSSFWKFI 1268

RESULT 7
Q933K0 PRELIMINARY; PRT; 1291 AA.
Q933K0:
AC Q933K0;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Type B cryptic neurotoxin.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
[1]
SEQUENCE FROM N.A.
STRAIN=593; AND 588;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
"Characterization of six type A strains of Clostridium botulinum that
contain type B toxin gene sequences."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF300466; AAL11499.1; -.
DR EMBL; AF300465; AAL11498.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase_M27; 1.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Neurotoxin.
SQ SEQUENCE 1291 AA; 150843 MW; 7AC1737B0FA5A151 CRC64;

Query Match 25.5%; Score 610; DB 2; Length 1291;
Best Local Similarity 33.9%; Pred. NO. 3.7e-24;
Matches 168; Conservative 84; Mismatches 154; Indels 90; Gaps 20;

2 TIPNIFSYTNSLLKDIINEYFNININDSKILSLQNRKNTLVDTSGYNAEVSDEGDVQLN 61
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
835 SIPFDLSTYNTNLTILIEIPKNYNSDILNLIILNRYDRDNKLDLSYGAKEVYDGVKLN 894

Qy 62 PIFPDKLSSGDRGKVIYTONENIVNSMYESFSISFWIRNKW---VSNL--PGY 115
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 895 D--KNQPKLTSSA--NSKIRVTQNIIFNSMFLDPSVFWIRIPKRYNDGIONTHNEY 950
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 116 TIIDSVKNNSGWSIGISNLFVTLTKQNEDEQSINFSYDINSNAPGY-NKWFFVTNNN 174
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 951 TIINCNNNSGWSIRGNMIWTLDINGKIKSVFEYSIKEDISEYINRWFVTITNN 1010
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 175 MGNMKYIYNGKLIDIKVKELTGNTFSITFEINKIPDGLTSDSDNNIN---MWIR 230
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1011 -SDNAKYIYNGKLESCHIDIRIEVIANDEANIEFKLD-----GNIDRTOFIWMK 1057
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 231 DFYIFAKELGKDNILFNSLOQTVNVKDYWGNDLRYNKEYYMNIDYLNRYM---AN 286
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1058 YFSFNTLSCQNIETIYKQSYSEVLKDFGPNPLMYKEYYFNAGNKNYSIKLKQSS 1117
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 287 SRQIVFTRNNN-----DFNEGYKIIIRKRGNT---NDTRVRGGDILYFDMTINK 336
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1118 VGEILLTRSKYQNASKYINRYDLYIGEKFIIR--KNSQSQSINDDIVRKEDYIYDFPNL 1176
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 337 AYNLFMKNETMYADNHSTEDIYALGREGTK---DINDNIIF---QIQPMNNTYIYAS 388
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1177 EWRVIM-----YKFKKEKELFLAPISDSDEFYNTIQIKKEYDEQPTVSC 1221
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```







Copied from 13910186 on 05-05-2004



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: November 7, 2002, 17:45:16 : Search time 1950 Seconds  
(without alignments)  
3737.420 Million cell updates/sec  
Title: US-09-910-186a-10  
Perfect score: 2391  
Sequence: 1 MTIPFNFSYTNNSLLKDI.....NYASLLESTHGWGFPVSE 450

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0  
Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
MODEL=frame+p2n.model -DEV=xlh  
-C/cgn2.1/USPTO.spool/US09910186/runat\_04112002\_111617\_20848/app\_query.fasta\_1.647  
DB=EST -QWMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
OUTFMT=ptc -NORM=ext -HAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
NO\_XLPXY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hct:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hct:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pla:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_nam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_oth:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127.5	5.3	870	17	AZ675343 ENTIG64TF
2	122.5	5.1	929	17	BH167076 ENTQX64TR
3	122	5.1	3707	14	C83838 C83838 Dict
4	119.5	5.0	917	17	AZ684830 ENTHZ58TF
5	119	5.0	649	9	AL297914 LP12358.5
6	117	4.9	575	17	AA549915 098m3.gm
7	116	4.9	562	13	BJ360994 BJ360994
8	115	4.8	641	12	BF297786 056PBC01
9	114.5	4.8	664	10	AW584247 N210227E
10	114	4.8	653	17	AA550238 135m3.gm
11	113.5	4.7	846	17	AZ689135 ENTGX61TF
12	112	4.7	549	13	BI13874 PFESTOa1
13	112	4.7	943	17	BH134547 ENTQI08TF
14	110	4.6	500	9	AU087833 AU087833
15	110	4.6	561	13	BM275413 PFESTOa6
16	110	4.6	690	13	BM162769 EST565303
17	110	4.6	945	17	AL404242 T7 end of
18	109.5	4.6	696	13	BM162617 EST565140
19	109.5	4.6	720	13	BM169137 EST571560
20	109.5	4.6	781	13	BM166487 EST569010
21	109	4.6	571	13	BM274132 PFESTOa6
22	108.5	4.5	597	14	BQ597280 PFESTOa3
23	108	4.5	638	14	BM815053 EST593147
24	108	4.5	754	17	AA550316 1457m3.gm
25	108	4.5	796	13	BM159196 EST561719
26	108	4.5	877	17	AZ683138 ENTML73TR
27	107.5	4.5	767	13	BM162128 EST564651
28	107	4.5	920	17	AZ547074 ENTEN03TF
29	107	4.5	568	13	BM275819 PFESTOa8
30	107	4.5	610	14	BQ595977 PFESTOa2
31	107	4.5	722	13	BM164970 EST567493
32	107	4.5	1225	9	AB012854 AB012854
33	106.5	4.5	662	17	AZ524207 228PBD12
34	106	4.4	561	13	BM274447 PFESTOa4
35	106	4.4	732	13	BM159156 EST561579
36	106	4.4	950	17	AZ693368 ENTJU13TR
37	105.5	4.4	688	12	BG604058 EST503148
38	105.5	4.4	694	12	BG604057 EST503147
39	105.5	4.4	698	13	BM163799 EST566322
40	105.5	4.4	764	13	BJ430749 BJ430749
41	105	4.4	1110	17	CNS06PWT
42	104.5	4.4	622	17	AZ526922 261PBC09
43	104.5	4.4	687	17	AZ526437 255PBE01
44	104.5	4.4	750	13	BM164094 EST566617
45	104.5	4.4	812	13	BM163086 EST565609

ALIGNMENTS

RESULT 1  
AZ675343  
LOCUS  
DEFINITION  
ENTIG64TF Entamoeba histolytica Sheared DNA linear GSS 14-DEC-2000  
ACCESSION  
AZ675343  
VERSION  
AZ675343.1 GI:11812609  
KEYWORDS  
GSS.  
SOURCE  
Entamoeba histolytica.  
ORGANISM  
Entamoeba histolytica.  
REFERENCE  
1 (bases 1 to 870)  
AUTHORS  
Loftus, B., Van Aken, S. and Fraser, C.  
TITLE  
Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library



JOURNAL  
COMMENT

Copied from 10910186 on 05-05-2004 by

Copied from 10910186 on 05-05-2004 by

1186 on 05-05-2004

86-05-05-2004

05-05 Ep 2004

05-05-2004

05-05-2004

-05-2004-

5 pp 2004

2004

Q4.

Qy

Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 77:7450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The  $v + i$  method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999).

BASE COUNT

Alignment scores:

Length:	0.000207	929
Matches:	132.50	65
Score:	37.74%	35
Conservative:	24.53%	92
Mismatches:	5.12%	74
Indels:	17	10
Gaps:	17	10

08-09-910-186A-10 (1-450) x BH167076 (1-929)

133	Ser	Asn	Phe	Leu	Val	Phe	Thr	Leu	Leu	Gln	Asn	Glu	Asp	Ser	Glu	Gln	Ser	Ile	Asn	Phe	151
																				152	
825	TCCGTT	TTTTT	ATAATA	TAATTCGTT	TAATAAATAAT															793	
153	Ser	Tyr	Asp	Ile	Ser	Asn	Asn	Ala	Pro	Gly	Tyr	Asn	Lys	Trp	Phe	Val	Thr	Val	Thr	172	
																			173		
792	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	796	
173	Asn	Asn	Met	Met	Gly	Asn	Met	Lys	Ile	Tyr	Ile	Asn	Gly	Lys	Leu	Ile	Asn	Phe	Thr	192	
																			193		
765	GTTT	AAATT	CAAG	GAATTTT	CTATATCGTT	GAAGGATGTTT	ATTATCATCAT	CAATTATT	706												
193	Val	Lys	Glu	Leu	Thr	Gly	Ile	Asn	Phe	Ser	Lys	Thr	Ile	Thr	Phe	Glu	Ile	Asn	Lys	212	
																			213		
705	CUGT	TTT	TTAT	TTC	CAT	TAATTTATTT	CAAGACTCT	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	708	
213	Pro	Asp	Thr	Gly	Leu	Ile	Thr	Ser	Asp	Ser	Asn	Ile	Asn	Met	Trp	Ile	Arg	Asp	Phe	232	
																			233		
648	CCA	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	608	
233	Tyr	Ile	Phe	Ala	Lys	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	237	
607	TATG	TGT	TGA	AGAA	CCAA	TAG	ATAG	ATAA	TCC	AAC	ATC	ATAC	ATAA	TA	AAAA	AAAA	AAAA	AAAA	AAAA	548	
238	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	248	
547	ATG	TCA	AT	AGC	AA	TAAT	TGC	AA	GA	AA	TAAT	TAAT	AT	GT	ATA	CCG	AA	TAC	CA	548	
																			549		
255	Asn	Val	Lys	Lys	Asp	Tyr	Trp	Gly	Asn	Asp	Leu	Arg	Tyr	Asn	Lys	-----	-----	-----	-----	271	
																			272		
487	ATG	AT	GTT	AT	G	AAG	AG	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	437	
272	Tyr	Met	Val	Asn	Ile	Asp	Tyr	Leu	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	290	
																			291		
436	TAC	ATA	TAT	ATA	AAG	TAT	TAT	CC	CAT	TAC	AA	TAC	AA	TAC	AA	TAC	AA	TAC	AA	377	
291	Val	Phe	Asn	Thr	Arg	Arg	Asn	Asn	Asn	Asp	Phe	Asn	Glu	Gly	-----	-----	-----	-----	-----	309	
																			310		
376	CA	AG	A	AG	A	AT	TAC	AC	G	CG	AA										

Qy	350	AspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluInThrIysAspIle	369
Db	220	-----TTTAAGAAAAACGAAGAATC	197
Qy	370	AsnAspAsnIleIle 374	
Db	196	GAAGAAAAAGACATT 182	
RESULT 3			
C83838			
LOCUS			
DEFINITION	C83838 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium		EST 28-APR-1999
ACCESSION	dictoideum CDNA clone SSA355, mRNA sequence.		
VERSION	C83838		
KEYWORDS	EST		
SOURCE	C83838.1 GI:2706770		
ORGANISM	Dictyostelium discoideum.		
REFERENCE	Eukaryote; Mycetozoa; Dictyostellida; Dictyostelium.		
AUTHORS	1 (bases 1 to 3707)		
TITLE	Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pi.M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai.H. and Tanaka,Y. The Dictyostelium developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development		
JOURNAL	DNA Res. 5 (6), 335-340 (1998)		
MEDLINE	99156227		
COMMENT	Contact: Hideko Urushihara Institute of Biological Sciences University of Tsukuba 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan Tel: 81-298-53-4664 Fax: 81-298-53-6614 Email: hideko@biol.tsukuba.ac.jp PROJECT = "Dictyostelium discoideum cDNA project in Japan" POLYA-No.		
FEATURES			
source	Location/Qualifiers		
	1..3707		
	/organism="Dictyostelium discoideum"		
	/strain="AX4"		
	/db_xref="taxon:44589"		
	/clone="SSA355"		
	/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"		
	/dev_stage="slug"		
BASE COUNT	1528 a 442 c 502 g 1235 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	0.00261	Length:	3707
Score:	122.00	Matches:	95
Percent Similarity:	35.37%	Conservative:	79
Best Local Similarity:	19.31%	Mismatches:	146
Query Match:	5.10%	Indels:	172
DB:	14	Gaps:	28
US-09-910-186A-10 (1-450); x C83838 (1-3707)			
Qy	48	TyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIle---	PheProPhe 66
Db	1649	TATTTTGATGAAATGATGATAAAAACAATATTAATTAATCACTCGCTATGTATCCACGT	1708
Qy	67	AspPheLysLeuGlySerSerGlyGluAspArgGly-	----- 78
Db	1709	GGATATAAATTAGGGCTTATTTTGATGAAGCAACCACCATTCATCTGTGATTTT	1768
Qy	79	-----LysValIleValThrGlnAsnGlu-	-----AsnIle 88
Db	1769	GATTTAAAATTTTCCAAGATGTTTAAACTCAAGATGACCTATTTCGCACCATGTTG	1828
Qy	89	ValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrp	108
Db	1829	ACCATATCTGT-----GATTTTGATTTAAAAATTTTCCAAAAGATGTTTAAACT	1879

109 valSerAsnLeuProGlyTyrThrIleileAspSerValIysAsnAsnSerGlyTyrPser 128  
 1880 CAAGATGACCTATTGGT-----AGTAATACTAGTAATTCa 1915  
 129 IleGlyIleIleSerAsnPhelLeuValPheThrLeuIysGln----- 142  
 1916 AGATTATTCAAAATCAACCAATGTTTCTCTCAAGTGCATATACTGGTGAAT 1975  
 143 ---AsnGluAspSerGluGlnSerIleAsnPhSerTyrAspIleSerAsnAsnAlaPro 161  
 1976 TGTAAATAGTAGTAATCAAGATATTCAAAATCAACTCAATGTTTCTCTATCAAGT 2035  
 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
 2036 GCATAT-----ACTGGTGGTAATTTTAAACAATACCATATC 2071  
 182 TyrIleAsnGlyLysLeuIleAspThr---IleIysValIysGluLeuThrGlyIleAsn 200  
 2072 TATGTCCTCAAAATAAGTTAATGATACGGTGTGAATATTCAACCTTTTCAATGGGTCA 2131  
 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220  
 2132 GCAACAATACCATTATCTATGTCGAAAATAAGTTAATGATACCGGTGCAAAATTCa 2191  
 221 -----AspSerAspAsnIle----- 225  
 2192 ACCTTTTCAATGGGCTCAGCAGATTGCTTTTATGTCACAGACAATTTTAAACCTTAT 2251  
 226 -----AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLys 242  
 2252 CAATTAATAATGAATGGGTGACTGGTTCATATTTTAAATGTTTATGTCGCCAAGA 2311  
 243 AspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn----- 255  
 2312 CAA-----TTTTTAACCTTTATCAATTAATAATGAATGGGTACTGGTTCCAT 2362  
 256 valValLysAsp-----TyrTrp----- 261  
 2363 ATTTTAAATAATAAATTTATTACTCTAGTAAAAGTGGTATCACTCAAAATGAATGTCOCA 2422  
 262 GlyAsnAspLeuArgTyr-----AsnLysGluTyrIle----- 272  
 2423 GGTACTGATCAAGATCTTATTAAACAATAAATTTATTACTCTAGTAAAAGTGGTATC 2482  
 273 -----MetValasnIleAspTyrLeuAsn 280  
 2483 ACTCAAAATGAATGCCAGGTACTGATCAAAATGTTCTTTATTAACAATGAATCTGCGAA 2542  
 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsn----- 297  
 2543 TCATTGTAATTTGTAATAATGAATAACAATTTATTATTACACAATGGTAAAATT 2602  
 298 AsnAsnAspPheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsn 317  
 2603 AAGAAAGAATCTGCGAATCAATTTGAAATTGTA-----AATPAATGAA 2644  
 318 AspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAla 337  
 2645 AARACA-----ATTATPAT---ATTACAACAATGGTAAAATT 2680  
 338 TyrAsnLeuPheMetLysAsn-----GluThrMetTyrAlaAspAsnHisSerThr 354  
 2681 AAGAAATAATGATGTTCAAAATTTAGTTTCATCAACACTTTAT---GATTTCATCTCGAT 2737  
 355 GluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIle 374  
 2738 AAA-----TCATTAGTGTGTTGTAATGTGCCCGGATGT---GAAATTTAGTT 2785  
 375 PheGlnIleGlnProMetAsnAsnThrTyrIle----- 385  
 2786 -----TCATCAACACTTTTATGATTCTCATACTCTGATAAATCATTAGGT 2827

QY	386	-----TyrAlSeGlnIllePheLysSerAsn	394
Db	2828	GTTTGTGAATGTCACCAGGTTTCTCTGGAAATTTATTGTAACTAATGTTTAATGGTACA	2887
QY	395	PheAsnGlyGluAsn-----IleSerGlyIle-----	403
Db	2888	GTTCAATGGGAATATGTAATCAACATGTTTAAAGATTTCTCTGGAATTTATTGTAAT	2947
QY	404	---CysSerIleGlyThrTyrArgPheArgLeuGly	414
Db	2948	CAATGTGTTAATGGTACACTTCAATGGGAAAATGGT	2983
RESULT 4			
LOCUS	AZ684830	917 bp DNA linear	GSS 14-DEC-2000
DEFINITION	ENTH256F Entamoeba histolytica Sheared DNA	Entamoeba histolytica genomic, DNA sequence.	
ACCESSION	AZ684830		
VERSION	AZ684830.1	GI:11821976	
KEYWORDS	GSS.		
SOURCE	Entamoeba histolytica.		
ORGANISM	Entamoeba histolytica.		
REFERENCE	Eukaryota; Entamoebidae; Entamoeba.		
AUTHORS	1 (bases 1 to 917)		
TITLE	Loftus,B., Van Aken,S. and Fraser,C. Determination of clone end sequences from Entamoeba histolytica		
JOURNAL	HMI:IMSS sheared DNA library		
COMMENT	Unpublished (2000) Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: bjloftus@tigr.org Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library		
FEATURES	Seq primer: M13-Forward Class: shotgun High quality sequence start: 24 High quality sequence stop: 871. Location/Qualifiers 1..917 /organism="Entamoeba histolytica" /strain="HMI:IMSS" /db_xref="taxon:5759" /clone_lib="Entamoeba histolytica Sheared DNA" /note="vector: PHOSI; Site_1: Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."		
BASE COUNT	390 a 104 c 153 g 270 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	0.000489	Length:	917
Score:	119.50	Matches:	77
Percent Similarity:	34.59%	Conservative:	61
Best Local Similarity:	19.30%	Mismatches:	116
Query Match:	5.00%	Indels:	145
DB:	17	Gaps:	17
US-09-910-186A-10 (1-450) x AZ684830 (1-917)			

US-09-910-186A-10 (1-450) x AZ684830 (1-917)



```

...      |||...      |||...      |||...      |||...
Db 168  ACGTTCATTCACCTCAACAAATAAATGGATGTGGTCAATATTCACGACCGGATAT 227
Qy 218  -----IleThrSerAspSerAsnIleAsnMetTrrp-----IleArgAspPheTyr 233
Db 228  TACGCTGTCACCTACTCCGACGACAT-----TGCATGCCATCCACACCTCTG 278
Qy 234  IlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyr 253
Db 279  ATCACACCAAC-----ValValLysAspTyrTrrpGlyAsnAspLeuArgTyrAsnLysGlu 270
Qy 254  ThrAsn-----ValValLysAspTyrTrrpGlyAsnAspLeuArgTyrAsnLysGlu 270
Db 309  AACATCGCCGCCAGATCGTGGATGATCTGTC--AACTGCCAGGCGCTGGATGTT 365
Qy 271  TyrTrrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIle 290
Db 366  ACCTACACACCTTACCTTGGATGATC--GAGTATTTGCAGCGGAGACCACTACATC 422
291  ValPheAsnThrArgAlaGAsnAsnAspPheAsnGluGlyTyrLysIleIleLys 310
423  CCTGGACATCGGCTTCAAC-----GGCTTCAACTACTTGACTATC 464
311  ArgIleArgGlyAsnThrAsnAspThrArgValArgGlyAspIleLeuTyrPheAsp 330
465  CGCCTC-----GGAAACGACACG-----GCCGATTCAATTACTATATC 503
331  MetThrIleAsnLysAlaTyrAsnLeuPheMetTyrAsnGlnThrMetTyrAlaAsp 350
504  CAGACCTCGACCAACAGGCTACATCAGCTGGCTTCATGAA-----GCATCG 554
351  AsnHisSerThrGluAspIleTyr 358
555  AATGATACAGCTTGGACATTTAC 578
AA549915 575 bp DNA linear GSS 05-DEC-2000
0988m3 gmbpFHB3.1, G. Roman Reddy Plasmodium falciparum genomic
clone 0988m, DNA sequence.
AA549915
AA549915.1 GI:2320167
GSS.
malaria parasite P. falciparum.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 575)
Dane, J.B., Arnot, D.E., Bourke, P., Chakrabarti, D., Christodoulou, Z.,
Coppel, R., Cowman, A., Craig, A., Fischer, K., Foster, J., Goodman, N.,
Hinterberg, K., Holder, A.A., Holt, D., Kemp, D., Lanzer, M., Lim, A.,
Newbold, C., Ravetch, J.V., Reddy, G.R., Rubio, J., Schuster, S.M., Su
X.-Z., Thompson, J.K., Vital, F., Wellens, T.E. and Werner, E.
Current status of the Plasmodium falciparum genome project
Mol. Biochem. Parasitol. 79, 1-12 (1996)
97001675
Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: T3
Class: shotgun.
Location/Qualifiers
1..575
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone="0988m"
/lab_host="E. coli XL1-Blue"
/note="vector: pBluescript SK(+); Genomic DNA, from
asynchronous blood stage parasites of the cloned Honduran

```

Copied from 10910186 on 05-05-2004

HB3 isolate cultured in vitro, was digested with mung bean nuclease in the presence of 30% formamide at 50°C (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the fragments were ligated to EcoR V-cleaved and dephosphorylated pBluescript SK(+). Recombinant plasmids transformed E. coli XL1-Blue.

BASE COUNT 286 a 45 c 68 g 175 t 1 Others  
ORIGIN

Alignment Scores: 0.000455 Length: 575  
Pred. No.: 117.00 Matches: 49  
Score: 36.73% Conservative: 23  
Percent Similarity: 25.00% Mismatches: 75  
Best Local Similarity: 4.89% Indels: 48  
Query Match: 17 Gaps: 8  
DB:

US-09-910-186a-10 (1-450) x AA549915 (1-575)

Qy 173 AsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly----- 185  
Db 48 AATAATATTATGGAAGAAAGAGGAATATTTAAATGCTGTTCTAAAGAACAAAGAA 107  
Qy 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205  
Db 108 AATTGGATGATATATATCCAAAATGAATATATGATGATAAACAATAAACAACA 167  
Qy 206 ThrPheGluIleAsn---LysIle-----ProAspThr 215  
Db 168 CAATATTTTAAATATGAATATTTTATTTACATTTTAAATATATATGCAAAACAGAA 227  
Qy 216 GlyLeuIleThrSerAsp-----SerAspAsnIleAsnMetTrrpIleArgAspPheTyr 233  
Db 228 TCATTAGATATAGATGATTTTAAATGATGAAAAAATATGCACTTGTTCGAATGTTA 287  
Qy 234 IlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyr 253  
Db 288 ATTAATAGAAACAACTTGTGATAAATGATTTGGAATATTTATTAATATGTTACAATTA 347  
Qy 254 ThrAsnValValLysAspTyrTrrpGlyAsnAspLeuArgTyrAsnLysGlnTyrMet 273  
Db 348 AAAAATAATGAAAAAGAAAAAATGAAAAAAGAAATATTAATCATATAATGAATATTTCAAG 407  
Qy 274 ValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsn 293  
Db 408 GAAATATATACCAT---GATAAATAGATATATACAAATCAATACCAATATGCTTAAT 464  
Qy 294 ThrArgArgAsnAsn---AsnAspPheAsnGluGlyTyrLysIleIleLysArgIle 312  
Db 465 ATGAATTATAACACTCCCAATAGGTACCGAGGAGGTAATAATATT----- 509  
Qy 313 ArgGlyAsnThrAsnAspThrArgValArgGlyLysPheAspMetThr 332  
Db 510 -----AATACTAATCCT----- 521  
Qy 333 IleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGlnThrMetTyr 348  
Db 522 -----AATGTGNTCAAGGAATGAAACAACTAT 551

RESULT 7  
BJ360994  
LOCUS  
DEFINITION BJ360994 Dictyostelium discoideum cDNA library, CF Dictyostelium  
discoideum cDNA clone ddc8p21 5', mRNA sequence.  
ACCESSION BJ360994  
VERSION BJ360994.1 GI:19260589  
KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum  
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
REFERENCE 1 (bases 1 to 562)

BJ360994 562 bp mRNA linear EST 07-MAR-2002  
Dictyostelium discoideum cDNA library, CF Dictyostelium  
discoideum cDNA clone ddc8p21 5', mRNA sequence.

SOURCE plasmodium berghei.  
ORGANISM Plasmodium berghei  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 641)  
AUTHORS Carlton, J. M.-R. and Dame, J. B.  
TITLE The Plasmodium vivax and P. berghei gene sequence tag projects  
JOURNAL Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)  
COMMENT Contact: Dame JB  
Department of Pathobiology, College of Veterinary Medicine  
University of Florida  
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611-0880, USA  
Tel: 352 332 4700  
Fax: 352 332 9704  
Email: damej@mail.vetmed.ufl.edu  
Seq primer: T3

```

FEATURES
    Location/Qualifiers
        1. .641
            /organism="Plasmodium berghei"
            /strain="ANKA Clone HP (gametocyte producer)"
            /db_xref="taxon:5821"
            /clone_lib="Pb cDNA #17, Tommaso Pace, Marta Ponzi, and
            Clara Frontali"

```

/lab\_host="Wistar rats"  
 /note="Vector: pBluescript II vector DNA, excised from  
 Lambda ZAP II.; Site1: EcoRI; Site2: XhoI; Total RNA was  
 extracted from asynchronous blood stage forms of the  
 cloned ANKA isolate of *P. berghei*, grown in Wistar rats to  
 30% parasitemia and 2-5% gametocytemia. Contaminating host  
 white cells had previously been removed and final host  
 cell contamination estimated to be approximately 5%.  
 PolyA+ RNA was extracted and reverse transcribed using an  
 oligo dT-XhoI primer (lambda ZAP II cDNA cloning kit,  
 Stratagene). Second strand cDNA was made following the  
 manufacturer's protocol. EcoRI adaptors were ligated to  
 the cDNA, and fragments were ligated into EcoRI/XhoI

```

BASE COUNT      332 a      68 c      85 g      156 t
ORIGIN
Alignment Scores:
Pred. No.:      0.000988
Score:          115.00
Length:         641
Matches:        62
Conservative:   31
Percent Similarity: 35.23%
Best Local Similarity: 23.48%
Mismatch:       83
Query Match:    18
Indels:         8
Gaps:           14
DB:             12
US-09-910-186A-10 (1-450) x BF297786 (1-641)

```

QY	139	ThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheserItyrAspIleSerAsn	158
		: : : : :     : : : : :     : : : : :	
Db	59	ACTAATATAAAGAAGAAACACCAAGAAACGACGAGATTCAAC	109
		: : : : :     : : : : :     : : : : :	
QY	159	AsnAlaProGlyItyrAsnLysItyrPhePheValThrValThrAsnAsnMetMetGlyAsn	178
		: : : : :     : : : : :     : : : : :	
Db	110	ATGTAATAACACGAATAATAGTGG	139
		: : : : :     : : : : :     : : : : :	
QY	179	MetLysIleItyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGly	198
		: : : : :     : : : : :     : : : : :	
Db	140	TTGAGA	145
		: : : : :     : : : : :     : : : : :	
QY	199	IleAsnPheserLysIthrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIle	218
		: : : : :     : : : : :     : : : : :	
Db	146	--AATGAATCTAATAAAATTTTTCAGAAATCAATATAGCTAATATGCAAT	193
		: : : : :     : : : : :     : : : : :	
QY	219	ThrSerAspSerAsnIle	234
		: : :       : : :     : : :     : : :	
Db	194	ATATATAAAGACGATAACTTACACATGATGAATAAT	229
		: : :       : : :     : : :     : : :	
QY	235	PheAlaLysGluLeuAspGlyLysAspIleAsnIle---LeuPheAsnSerLeuGlnItyr	253
		: : :       : : :     : : :     : : :	
Db	230	-----AGGAGATTTGGAT---AAACATGGTAAATTAGCAAAATTTCCGACAAATATTATT	280
		: : :       : : :     : : :     : : :	

[illegible]

Qy 254 ThrAsnValVallyAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMet 273  
 Db 281 AACACGAAATAAATATGGG---AATAAATACACACGCAATTAACAAATCTAATATAAT 337  
 Qy 274 ValAsnLeuAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsn 293  
 Db 338 AAAATATGATATATTCAGAAATCAAAATATAAATTTACAGAAATTTAAATAATAC 397  
 Qy 294 ThrArgArgAsnAsnAspPheAsnGluGlyTyrTyrLysIleIleLysArgIleArg 313  
 Db 398 TACATAAAATACAAACAAATTTAATCATGAATAC----- 433  
 Qy 314 GlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIle 333  
 Db 434 GGGATCAATACAAACAAATCGGATAGGA-----GATGAACGAAC 475  
 Qy 334 AsnAsnLysAlaTyrAsn---LeuPheMetLysAsnGluThrMetTyrAlaAsnAsnHis 352  
 Db 476 AATAATAATAGTGATAACATATTGTCATGAACAATCTAAGCAACTGCAATAGT--- 532  
 Qy 353 SerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsn 372  
 Db 533 ----- 547  
 Qy 373 IleIlePheGlnIleGlnPrometAsnAsnThrTyrTyrAlaSerGlnIlePheLys 392  
 Db 548 -----AATCCCAAAATGGGAAATACACGATTCACCAATTTGCTGAT 592  
 Qy 393 SerAsnPheAsn 396  
 Db 593 GGCAATTTCAAT 604

AW584247 664 bp mRNA linear EST 07-SEP-2000  
 N210227e MHAM Medicago truncatula/Glomis versiforme mixed EST  
 library cDNA clone MHAM-1F21, mRNA sequence.

AW584247 GI:7261301  
 Medicago truncatula/Glomis versiforme mixed EST library.  
 Medicago truncatula/Glomis versiforme mixed EST library  
 Eukaryotes: mixed EST libraries.  
 1 (bases 1 to 664)  
 Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,  
 Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
 ESTs from roots of Medicago truncatula after colonization with  
 Glomis versiforme  
 Unpublished (2000)  
 Contact: Harrison M.J.  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73401  
 Tel: 580-223-5810  
 Fax: 580-221-7380  
 Email: mjharrison@ncbnc.org  
 Other name: MHAM-1c-11; Date: 3/14/00; Updated to the Database of  
 Expressed Sequence Tags (dbEST) on 04/27/00; More information is  
 available at 'http://chryslr.tamu.edu/medicago'.  
 Seq primer: T3.  
 Location/Qualifiers  
 1. 664  
 /organism="Medicago truncatula/Glomis versiforme mixed EST  
 library"  
 /cultivar="Medicago truncatula genotype A17"  
 /db\_xref="taxon:119092"  
 /clone="MHAM-1F21"  
 /clone\_lib="MHAM"  
 /tissue\_type="roots colonized with Glomis versiforme"  
 /dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomis versiforme. The library was  
 made from a mixture of RNA from each of these stages."

2004-09-03 10:10:10 109410186a-10

/lab\_host="E. coli strain XL0LR"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; cDNA was prepared from polyA+ enriched RNA from  
 roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomis versiforme. The cDNA was  
 directionally ligated into the Unizap XR vector from  
 StrataGene and packaged using Gigapack III Gold packaging  
 extracts. Plasmids containing cDNA inserts were excised  
 from the recombinant lambda-zap phage using Ex-assist  
 helper phage and propagated in XL0LR cells."  
 BASE COUNT 297 a 142 c 96 g 129 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.00122 Length: 664  
 Score: 114.50 Matches: 73  
 Percent Similarity: 31.61% Conservatives: 37  
 Best Local Similarity: 20.98% Mismatches: 95  
 Query Match: 4.79% Indels: 143  
 DB: 10 Gaps: 19

US-09-910-186a-10 (1-450) x AW584247 (1-664)

Qy 4 ProPheAsnIlePhe---SerTyrThrAsnAsnSerLeuLeuLysAspIle----- 19  
 Db 4 CCAGCCACACACTTACCAATCATATGAAAAAAGTCTCTCACACGATGAATACAAACAAC 63  
 Qy 20 ---IleAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArg 38  
 Db 64 AAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 102  
 Qy 39 LysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspVal 58  
 Db 103 -----AACAGTACCCGAATGAAGTATGATGACACA 132  
 Qy 59 GlnLeuAsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGly 78  
 Db 133 AAATACACACACA-----TCTTACAGCAGCAGCAGCAACAAT----- 171  
 Qy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98  
 Db 172 -----AACTACGACACAAACAGTATCTTCTAGTCTGATGACTTACG 216  
 Qy 99 IleSerPheThrIleArgIleAsnLysTyrPheValSerAsnLeuPro---GlyTyrThr 116  
 Db 217 -----ACAAA-----AAGTCCAGAAAGAGATATAAC 246  
 Qy 117 IleIleAspSerValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeu 136  
 Db 247 TCGAACGAAATCAGAAACAAACAAC-----AACGATAACACAG 285  
 Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 286 TATTCTACACACACAAACAGGATCTCTAGCAACACAAAGTTTACAGAGAA--- 342  
 Qy 157 SerAsnAlaProGlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMet 176  
 Db 343 -----GGATACAAC-----TCCATGGAAAACCGGACAAAC 372  
 Qy 177 GlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeu 196  
 Db 373 AACACACAAACAGTATTCTTACACACACAAACAACAAAGAT----- 411  
 Qy 197 ThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGly 216  
 Db 412 -----TCTTCTAGCAACACA-----AAGTCCAGAAAGAGGT 444  
 Qy 217 LeuIleThrSerAspSerAspAsnIleAsnMetIleArgAspPheThrIlePheAla 236  
 Db 444 ----- 444  
 Qy 237 LysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnVal 256  
 Db ----- 256

Db 445 -----TACAACCTCAATGCAAAATCAGAAC--- 468  
 QY 257 ValysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIle 276  
 Db 469 -----ACACAACTAC----- 480  
 QY 277 AspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArg--- 295  
 Db 481 -----GAGAAGTATAACTACACAAAC-----AAAGTTGCTGTTAATGACAAATAC 525  
 QY 296 -----ArgAsnAsnAspPheAsnGluGlyTyrLysIleIleLysArgIleArg 313  
 Db 526 AGCTTCAAAGTAAACACCAATACAAACATGCTGAG-----AGG 564  
 QY 314 GlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIle 333  
 Db 565 CAAGGAATGAGTCACACGAGAGTATGGAAGCTGGAAATATTTTATGATGCTTAACTCT 624  
 QY 334 AsnAsnLysAlaTyrAsnLeuPhe 341  
 Db 625 GAGGAGAAATATACCCACATTC 648  
 RESULT 10  
 LOCUS AA550238/c 653 bp DNA linear GSS 05-DEC-2000  
 DEFINITION 1358m3 gmbpfHB3.1, G. Roman Reddy Plasmodium falciparum genomic  
 accession AA550238  
 VERSION AA550238.1 GI:2320490  
 KEYWORDS GSS.  
 SOURCE malaria parasite P. falciparum.  
 ORGANISM Plasmodium falciparum  
 REFERENCE 1 (bases 1 to 653)  
 AUTHORS Dame, J.B., Arnott, D.E., Bourke, P., Chakrabarti, D., Christodoulou, Z.,  
 Coppel, R., Cowman, A., Craig, A., Fischer, K., Foster, J., Goodman, N.,  
 Hinterberger, K., Holder, A.A., Holt, D., Kemp, D., Lanzier, M., Lim, A.,  
 Newbold, C., Ravetch, J.V., Reddy, G.R., Rubio, J., Schuster, S.M., Su  
 X.-Z., Thompson, J.K., Vital, F., Wellens, T.E. and Werner, E.  
 Current status of the Plasmodium falciparum genome project  
 Mol. Biochem. Parasitol. 79, 1-12 (1996)  
 CONTACT: Dame JB  
 Dept. of Pathobiology, College of Veterinary Medicine  
 University of Florida  
 2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA  
 Tel: 352 392 4700  
 Fax: 352 392 9704  
 Email: damej@mail.ufl.edu  
 Seq primer: T3  
 Class: shotgun.

## FEATURES

## source

1. .653  
 /organism="Plasmodium falciparum"  
 /db\_xref="taxon:5833"  
 /clone="1358m"  
 /lab\_host="E. coli XL1-Blue"  
 /note="Vector: pBluescript SK(+); Genomic DNA, from  
 asynchronous blood stage parasites of the cloned Honduran  
 HB3 isolate cultured in vitro, was digested with HincII  
 nuclease in the presence of 30% formamide at 50°C (Vernick  
 & K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic  
 Acids Research 16:5883-5896). The ends of the fragments  
 were ligated to EcoRV-cleaved and dephosphorylated  
 pBluescript SK(+). Recombinant plasmids transformed E.  
 coli XL1-Blue."

BASE COUNT 193 a 84 c 50 g 326 t  
 ORIGIN

Alignment Scores:

Pred. No.: 0.00137 Length: 653

Score: 114.00 Matches: 49  
 Percent Similarity: 39.71% Conservative: 32  
 Best Local Similarity: 24.02% Mismatches: 81  
 Query Match: 4.77% Indels: 42  
 DB: 17 Gaps: 10  
 US-09-910-186a-10 (1-450) x AA550238 (1-653)  
 QY 253 TTTThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyr 272  
 Db 645 TATAATGACAAATAAG-----AAAAACAAAAAATAATAGTATTATGAA 598  
 QY 273 MetValAsnIleAspTyrLeuAsnArgTyrMetTyrAla----- 285  
 Db 597 ATCAATAAATACATATATGAATGAACATGGATATACAGATATTGAAATGACGTTTA 538  
 QY 286 AsnSerArgGlnIleValPheAsnThrArgAsnAsnAspPheAsnGluGlyTyr 305  
 Db 537 AATAAAAAAATAACGATTGAATGAGAGCTCGTACCAATACATTGGATGAT----- 484  
 QY 306 LysIleIleLysArgIleArgGlyAsnThrAsnAspThr----- 319  
 Db 483 -----ATTATTGTAAGTCATCATGGAATAGTTATGATAAATAATAATACAGTAAACAT 430  
 QY 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339  
 Db 429 AATAGAGAAAGCAATCATATAATGAGATGAGAAAGAAACAGAAATATAAA----- 379  
 QY 340 LeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAla 359  
 Db 378 -----AAAAAACACGTTATTGTAGATGA-----AAAGATATGGAAGGA 337  
 QY 360 IleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPro 379  
 Db 336 ATAGGAAAGAAAGAAATAAATAATAGTAATATATATTTTAT----- 289  
 QY 380 MetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGlu--- 398  
 Db 288 ---AATAACTCATATAGTATAATAATAATAGTCTCTATAGTAATAATAAATGATATA 232  
 QY 399 ---AsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyAspTyr 417  
 Db 231 TATAGTGTGATTAATGATCATCTGTATTAATAACAATAATGATATCATGAGGTCCCAAGC 172  
 QY 418 TTTArgHis-----AsnTyrLeuValProThrValLysGlnGly-----AsnTyr 432  
 Db 171 TAGTCTCATGCTACTTATTATAAACAAGTAAATGATATATATCAAGGCTACCTAATAT 112  
 QY 433 AlaSerLeuLeu 436  
 Db 111 AATAACATGATG 100

## RESULT 11

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

AZ689135 846 bp DNA linear GSS 14-DEC-2000  
 ENTX61TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 genomic, DNA sequence.

AZ689135  
 AZ689135.1 GI:11826281

Entamoeba histolytica.  
 Entamoeba histolytica.

Entamoeba histolytica.  
 Entamoebidae; Entamoeba.

1 (bases 1 to 846)  
 Loftus, B., Van Aken, S. and Fraser, C.

Determination of clone end sequences from Entamoeba histolytica  
 HMI:IMSS sheared DNA library

Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208





```

source          1. 549
/organism="Plasmodium falciparum"
/db_xref="taxon:9833"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/Note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
XhoI; Library was constructed by Debopam Chakrabarti.
Total RNA samples were isolated from mixed stage
saponin(0.1%) lysed P. falciparum 3D7 infected
erythrocytes by the acidic guanidinium-phenol chloroform
method. The poly A+ RNA was isolated by the polyAI-tract
mRNA isolation system (Promega, WI) using streptavidin
magnisphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
into EcoRI and XhoI sites of 1 zapII vector using the Zap
cDNA synthesis kit (Stratagene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the Exassist helper phage
(Stratagene), the phagemids were precipitated with PEG
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."
BASE COUNT      278 a  51 c  54 g  166 t
ORIGIN
Alignment Scores:
Score:          0.00182      Length:          549
Percent Similarity: 112.00      Matches:          466
Best Local Similarity: 41.01%    Conservative:    27
Query Match:    25.84%    Mismatches:      63
Indels:         42
Gaps:           13
US-09-910-186a-10 (1-450) x BI813874 (1-549)
222 SerAspAsnMetTrpIleArgAspPheThrIlePheAlaLys----- 237
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
60 AGTCAAAATATTTCAAATATAATATGATGATAAAGAAATGATAAATAATCAACTA 119
238 GluLeuAspGlyLeuAspIleLeuPheAsnSerLeuGluTyrThrAsnValVal 257
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 AAAATGTCGAGAAAGATATTATGATATATCATCCATTGGATAAATAATGATANA 179
258 LysAspTyrTrpGlyAsnAspLeuArg----- 266
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 AAAAATTTTATGCAATGATATGTCACCACTCAAAATAGTTCGAAAATTTCTCTTTT 239
267 -----TyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsn 280
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
240 GAAAAACCAATGGAATAATAATCAAACTATAATCTTGTAAAT---AAATATTTAAAT 296
281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAlaAsnAsp 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
297 CATGAAAGATCAAAACATAAATAGTGTACATAATTAAATGATAATAATAATAAAT 356
301 PheAsnGluGlyTyrLysIleleleleLysArgIleArgGlyAsnThrAspThrArg 320
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
357 AATAATAAAGAGAATAATATATATATATATATATATATATATATATATATATATAT 413
321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAspLysAlaTyrAsnLeu 340
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
414 -----GATATATATAT-----ACATATAATAATTCATCAATAT 449
341 Phe-----MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyr 358
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
450 TTTGATGAAATTCGAAAGAAATGGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 503
359 AlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln 376
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
504 -----AAAAAANAACATCAAGAACACAGACTACTTATACAA 536
RESULT 13
BH134547/c
LOCUS

```

ORIGIN

Alignment Scores:  
Score: 0.00279 Length: 500  
Matches: 110.00  
Percent Similarity: 39.61% Conservative: 17  
Best Local Similarity: 28.57% Mismatches: 59  
Query Match: 4.60% Indels: 34  
Gaps: 5

US-09-910-186A-10 (1-450) x AU087833 (1-500)

QY 234 lIePheAlarLyLeuAspGLyLysAspIlleAsnIleLeuPheAsnSerLeuGlntyr 253  
||||| :||| :||| :||| :||| :|||  
Db 94 ATATTGGTTTGACACATCAGGTGAGAATAATATTATAACAACGATAAGTATATA 153  
QY 254 ThrAsnValLysAspTyTrpGLyVAscAspLeuArgTyAsnLysGLuTyTrMet 273  
||||| :||| :||| :||| :||| :|||  
Db 154 AATAA-----GGAATAGG----- 168  
QY 274 ValAsnIleAspTyTrLeuAsnArgTyMetTyTrAla-----Asser 287  
||||| :||| :||| :||| :||| :|||  
Db 169 --AATGTAGATGAGGAATGAAAAGATCTTAGTTAATAATGAAGGATATAATAA 225  
QY 288 ArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGlu----- 303  
:||| :||| :||| :||| :||| :|||  
Db 226 AGCAATTGTTAAAAATGATCATATATTAATATGATTTAATAAAAAATATATAT 285  
QY 304 -----GlyTyThyllelleLysArgIleArgGlyAsnThrAsnsp 318  
:||| :||| :||| :||| :||| :|||  
Db 286 GACAAACATAATAGTAATGAAAAAATATATCAAGTAATAAATNCAAATATAAATGAT 345  
QY 319 ThrArgValArgGlyGlyAspIlleLeuTyPheAspMetThrIleAsnAsnLysAlatyr 338  
:||| :||| :||| :||| :||| :|||  
Db 346 ATAATAAATATTATGTAGTAAATAATCTTGATCGAAAAATAATTTAATAAGATAGA 405  
QY 339 AsnLeuPheMetLysAsnGluThrMetTyTrAlaAspAsnHisSerThrGluaspIleTy 358  
:||| :||| :||| :||| :||| :|||  
Db 406 TCCAAATTAGAGATTAGAGACATTC-----AACGAGATGCACAAGATTAAATAAT 456  
QY 359 AlaileGlyLeuArgGluGlnThrLysAspIlleAsnAspAsn 372  
:||| :||| :||| :||| :||| :|||  
Db 457 GCATGTCATATATGAATAAATGATATATATAATGATAAT 498

RESULT 15  
BM275413 561 bp mRNA linear EST 20-DEC-2001  
PFESTceae8e04.y1 Plasmodium falciparum 3D7 gametocyte cDNA library  
Plasmodium falciparum 3D7 CDNA 5', mRNA sequence.

BM275413  
BM275413.1 GI:17968741  
EST.  
Plasmodium falciparum 3D7.  
Plasmodium falciparum 3D7.  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
1 (bases 1 to 561)  
Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,  
Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,  
Bowers,Y., Gibbons,M., Bitter,E., Bennett,J., Jentes,E., Ronko,I.,  
Teagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,  
Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,  
Waterston,R., Wilson,K. and Sibley,D.  
Washo Plasmodium EST Project  
Unpublished (2001)  
Contact: L. David Sibley  
Washo Plasmodium EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library was constructed by R. Haywood. DNA sequencing by:  
Washington University Genome Sequencing Center For information on



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 11:55:50 ; Search time 2773 Seconds  
(without alignments)  
14388.729 Million cell updates/sec

Title: US-09-910-186A-9

Perfect score: 1371

Sequence: 1 gaattcagatgaccatccc.....tctccgagtaatggaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_ov.\*  
5: gb\_ov.\*  
6: gb\_ov.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pi.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rod.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgt\_hum.\*  
40: em\_hgt\_mus.\*  
41: em\_hgt\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	665.2	48.5	4479	7	CBCPHGC1	X53751 Clostridium
2	665.2	48.5	4592	7	CSTCITOX	D90210 Bacterioph
3	665.2	48.5	4712	1	CBCPHCNLM	X71126 C.baculum
4	665.2	48.5	9613	1	CBCTOX	X62389 Botulinum b
5	665.2	48.5	9689	1	CBP1NCTC1	X66433 Clostridium
6	665.2	48.5	11747	1	AB061780	AB061780 Clostridi
7	665.2	48.5	12297	7	CBCBONT	X72793 Clostridium
8	449.8	32.8	11631	1	AB037920	AB037920 Clostridi
9	448.2	32.7	4049	1	CLOTNCO	D38442 Clostridium
10	189	13.8	151705	2	AC117900	AC117900 Rattus no
11	188.4	13.7	176351	2	AC117007	AC117007 Rattus no
12	183	13.3	169163	2	AC115666	AC115666 Rattus no
13	180.2	13.1	163034	2	AC099432	AC099432 Rattus no
14	170.2	12.4	165337	2	AC096212	AC096212 Rattus no
15	169.4	12.4	1402	6	AR000030	AR000030 Sequence
16	169.4	12.4	1402	6	AR169141	AR169141 Sequence
17	169.4	12.4	1402	6	AX036246	AX036246 Sequence
18	168.8	12.3	1330	6	AR000029	AR000029 Sequence
19	168.8	12.3	1330	6	AR169140	AR169140 Sequence
20	168.8	12.3	1330	6	AX036243	AX036243 Sequence
21	168.8	12.3	1338	12	XXU22962	U22962 Synthetic b
22	168.8	12.3	97683	2	AC116548	AC116548 Dictyoste
23	165.6	12.1	236942	2	AC096320	AC096320 Rattus no
24	164.6	12.0	256774	2	AC116964	AC116964 Dictyoste
25	160.2	11.7	149228	2	AC113774	AC113774 Rattus no
26	158.8	11.6	115758	9	AC104634	AC104634 Homo sapi
27	157	11.5	108476	2	AC094507	AC094507 Rattus no
28	156.8	11.4	158063	9	AP001046	AP001046 Homo sapi
29	156.8	11.4	340000	9	AP001751	AP001751 Homo sapi
30	155.4	11.3	173540	2	AC119697	AC119697 Rattus no
31	155.2	11.3	131346	2	AC119558	AC119558 Rattus no
32	154.6	11.3	169966	2	AC126213	AC126213 Rattus no
33	153.2	11.2	1299	12	AF251281	AF251281 Synthetic
34	153	11.2	7921	3	AF153362	AF153362 Dictyoste
35	151.6	11.1	1313	6	A58946	A58946 Sequence 6
36	151.6	11.1	39369	2	AC115681	AC115681 Dictyoste
37	150.6	11.0	186278	9	AC079176	AC079176 Homo sapi
38	147.6	10.8	229502	2	AC126871	AC126871 Rattus no
39	145.8	10.6	106177	2	AC096493	AC096493 Rattus no
40	145.8	10.6	125026	10	AC091616	AC091616 Rattus no
41	144.2	10.5	156533	2	AC117070	AC117070 Dictyoste
42	143.8	10.5	157141	3	AC016445	AC016445 Drosophil
43	136.6	10.0	147671	2	AC105625	AC105625 Rattus no
44	136	9.9	5163	6	AR097042	AR097042 Sequence
45	136	9.9	5318	6	AR097041	AR097041 Sequence

ALIGNMENTS

RESULT 1	CBCPHGC1	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
LOCUS	X53751	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
DEFINITION	X53751	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
ACCESSION	X53751	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
VERSION	X53751.1	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
KEYWORDS	Cl/neurotoxin; neurotoxin; toxin.	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
SOURCE	Clostridium botulinum C	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
ORGANISM	Clostridium botulinum C	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
REFERENCE	1 (bases 1 to 4479)	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
AUTHORS	Popoff M.R.	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
TITLE	Direct Submission	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
JOURNAL	Submitted (03-JUL-1990)	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993
	25 rue du Docteur Roux, 75724 Paris Cedex 15, France	Clostridium botulinum C	4479 bp	DNA	linear	PHG 12-SEP-1993

Pred. No. is the number of results predicted by chance to have a

```

REMARK revised by [3]
REFERENCE 2 (bases 1 to 4479)
AUTHORS Hauser, D., Eklund, M.W., Kurazono, H., Binz, T., Niemann, H.,
Gill, D.M., Boquet, P. and Popoff, M.R.
TITLE Nucleotide sequence of Clostridium botulinum C1 neurotoxin
JOURNAL Nucleic Acids Res. 18 (16), 4924 (1990)
MEDLINE 90370487
PUBMED 2204031
REMARK revised by [3]
REFERENCE 3 (bases 1 to 4479)
AUTHORS Popoff, M.R.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1990)
FEATURES Location/Qualifiers
source
1..4479
/organism="Clostridium botulinum C phase"
/strain="phage from 468C botulinum C. strain"
/db_xref="taxon:12448"
198..203
/note="pot. ribosome binding site"
214..4089
/note="C1 neurotoxin (AA 1-1291)"
/codon_start=1
/transl_table=1
/protein_id="CAA37780.1"
/db_xref="GI:14906"
/db_xref="SWISS-PROT:P18640"
/translation="MPEITNNYSDVDNKNILYLDTHLNTLANEPEKAFRTICNLW
VIDPERSRNSPLNKPRTVSPKSGYDPPNVLSDSKDPLKEIKLEKINSREI
GSEILYRLSDIPFPNNVPTINTDFDVFNSVDKTRQGNWYKGSINPVLITG
FCNIIDPSTFSLKNTNFAAGFGLSIIISPRFMTYSNATNDVGGREKSE
PMDPILILHMLHNAHMLYGLAIAPQDTSSTNIFSYQYNKLEYAEIYAFGP
TIDLPKARKYFEKALDYSLAKRLNITTPNPFNKYIGYKQKLRKYRFVY
ESGEVTVNRKVFVYELNLTQFTEFNKYIYNVONRKIYLSNVYPTVANIIDDVY
YDQNGFNLPKSNLVLFGONLSRNPALRKYNPENMLYLFKFCRAIDGRSLYNT
LDCELLYNTDLPLFGLSDVKTDLFLKDLNEETEVYVYDNDVSDVILSKNTSE
HGQDLVLSIDSESLIPGNQVYDNRQNTQNDVLYNLYLESQKSDNVEDFTFR
SIEELDNDAKVTYFPTLANKNVAGVGGFLMWANDVEDFTNILLKDLIDKISD
VSAIPIYGPALNISVSRGNETFAFVGTILLEAPFEFTTIPNILLKDLIDKISD
RNFIIKIDNCKEIRKWKDSYENMGITWLSRIITFPNNTSYQMDSLNVOAGAIA
KIDLEYKYSKGSQENKQVNSLDKISEAMNINFKIFRECSVYLFKRWLPR
VIDELEFRNTKALINILDSHNIILGEVDKLAQVNSFPONTIPNIEFSYNNLS
LKDIINEYFNNDKLSLONKNTLVDTSYNAEVEEDGVOLNLPFPDFKLGSS
GERGKVTYVONENIYVNSWYSEFSISWIRINKVWNLPGYTIIDSVKNNSHSGI
ISNPLFTLQNDSEQSINFSDISNAPGPKWFFVYVNNMGMNKITKINGKLDIL
TIKVLGTGINSKTIIFINKIPDGLTIDSDNINMWIRDFIFAKELGDKDNLIL
FNSLQTYNVYDNGDLRYNKEYVYNIDLYNRYNANRQIVFENTERNNDNFEGY
KIIIKRIGNTDTRVRGGDILYFDMTINKKAYNLFMKNETMYADNHSFEDIYALGR
EQDKINDNTIFQIOPMNTYVYASQLEKSNFNGENISGICSGIYRFRILGDDNTRHN
YLVPYKQGNVYASLLESTHMGFVYVSE"
BASE COUNT 1817 a 455 c 693 g 1514 t
BEGIN
Query Match 48.5%; Score 665.2; DB 7; Length 4479;
Best Local Similarity 68.3%; Pred. No. 1.6e-92;
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;
QY 13 ACCATCCATTCAACATCTCTCTACACCAACAACTCCCTGTGAGGACATCATCAC 72
Db 2740 ACAATACCCCTTTAATATTTTTCATACATACTAATAAATCTTTTATAAAGATATAATTAAT 2799
QY 73 GAGTACTTCAACACATCAAGACTCCCAAGATCTGCCCTCGAGAACCGTAAAGACACC 132
Db 2800 GAATATTTCAATATATTAATGATTCAAAAATTTTGAGCCTACAAACAGAAAAATACT 2859
QY 133 TTGTCGACACCTCCGGTTACAACCGAGGTCTCCGAGGAGGGTGACGTCCAGCTGAAC 192
Db 2860 TTATGGGATACATCAGGATATAATGCAAGAGTGAAGTGAAGAGCGATGTTTCAGCTTAAT 2919
QY 193 CCAATCTTCCATCGACTTCAACTGGTCTCTCGGTGAGGACAGAGGTAAAGGTCATC 252
Db 2920 CCAATATTTCCATTTGACTTTTAATATTAGTAGTTCAGGGGAGGATAGAGGTAAAGTTATA 2979

```

Db 4060 CATGGGGTTTGCTAGCTGAAGTAATAA 4089  
 || ||||| ||||| ||||| ||||| ||||| |||||

## RESULT 2

CSTCITOX 4592 bp DNA linear PHG 23-JUN-1999  
 LOCUS Bacteriophage c-st (from C. botulinum) C1-tox gene for botulinum C1  
 DEFINITION neurotoxin.  
 ACCESSION D90210.1 GI:2117780  
 VERSION 1  
 KEYWORDS botulinum toxin; neurotoxin.  
 SOURCE Bacteriophage c-st (from C. botulinum type C-Stockholm) DNA.  
 ORGANISM Bacteriophage c-st  
 Viruses.

REFERENCE 1 (bases 1 to 4592)  
 Kimura.K., Fujii.N., Tsuzuki.K., Murakami.T., Indoh.T., Yokosawa.N.  
 Yokusawa.N., Takeshi.K., Syuto.B. and Oguma.K.  
 TITLE The complete nucleotide sequence of the gene coding for botulinum  
 type C1 toxin in the C-ST phage genome  
 JOURNAL Biochem. Biophys. Res. Commun. 171 (3), 1304-1311 (1990)  
 MEDLINE 91024998  
 REFERENCE 2 (bases 36 to 2143)  
 Kimura.K., Fujii.N., Tsuzuki.K., Murakami.T., Indoh.T., Yokosawa.N.  
 and Oguma.K.  
 TITLE Cloning of the structural gene for Clostridium botulinum type C1  
 toxin and whole nucleotide sequence of its light chain component  
 JOURNAL Appl. Environ. Microbiol. 57 (4), 1168-1172 (1991)  
 MEDLINE 91282468  
 COMMENT These data kindly submitted in computer readable form by: Kouichi  
 Kimura  
 Department of Microbiology  
 Sapporo Medical College  
 1 South, 17 West  
 Sapporo 060  
 Japan  
 E-mail: nfujii@niguts.nig.ac.jp  
 Phone: 011-611-2111 x2334  
 Fax: 011-612-5861.

Location/Qualifiers

1. 4592  
 /organism="Bacteriophage c-st"  
 /db\_xref="taxon:12336"  
 255..260  
 293..298  
 305..310  
 /note="SD sequence"  
 321..4196  
 /note="ORF for C1 toxin"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAAL4235.1"  
 /db\_xref="GI:2117781"

## FEATURES

source  
 -35\_signal  
 -10\_signal  
 RBS  
 CDS  
 Translation="MPTINNFNSDPVDNKNILYLDTHLNLANEPEKAFRTGNITW  
 VTPDRFSNPNLNKPRVTPSPKSGYVDNVLSTDSKDTFLKEIKLPRKNSRI  
 GELLYRLSTDIPEPNNNTPIINFDVDNVDKTRGNWVKTSINPSVILTG  
 PRENIDETSTFKLINTPAQEGFALSIISIPRNLIYSNATNDVGEFSEKSE  
 FCMDFILMLHNLHNLGIALPNDPDISSVSNIFYSQYNKLEYIAEIGV  
 ESSGEVYNNKFEVLYNELQIFTEFNKYLNNQKRYLSNVYTPVTANILDDNV  
 YIQGNFIPKSNLNLPMGQNLNRPALRKVNENMLYLFKFKRAIDGRSLYKNT  
 LDCRELLVNTDLPIGIDSDVTDIFLRKDINEETEVYYPDNVSDVQVILSKNTSE  
 HQCLLDYSDISESEILPGENQVFNDRNTONVYLSNXYLESQKLSNVDVETPR  
 STEEALDSKAYVYFPTLANKNVAGVGGGLFLMWANDVVEFTNLRKDLTKSD  
 VSAIPYIGPALNLSNVRNGNFTFAVIGVTILLDEAFPEFTIPALGCAVYSKQVE  
 RNEILTKIDNCELRKIKWDSYEMWGTWLSRIITQFNISYQMYDSLANQAGALKA  
 KIDLEYKSGSDKNITKSNQVNLKNSLDVKISEAMNNINKPFIRESVYLFKNMPLK  
 VIDENEFNRTKAKLNLIDSHNIIIVGEVDKLRKAKVNSFQNTIPFNIFSYTNKL  
 LKIDINEYFNNDNLKSLSLQNRKNTLVDTSGNAYESEGVDOLNPIFPDFEKLGS  
 GDRGKVLTONENTVYNSMESIESIPWIRINKVWSNLPCTIIDSVKNNWSGI  
 ISNLFVFLKQNEDESEQINSYDISNAPGYKWPVYVYNNMGNKYYINGKLID  
 TIKVKELTGINSKFTFEINKPDTGLITSDSNINKNWIRDFYIFAKELDGDKNIL  
 FNSLQYTNVVKDYGNDLRYNKNKYMYNIDYLNRYMYANSRQIVFNTFRNNNDPNEG

Query Match 48.5%; Score 665.2; DB 7; Length 4592;  
 Best Local Similarity 68.3%; Pred. No. 1.5e-92;  
 Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;  
 BASE COUNT 1869 a 460 c 714 g 1549 t  
 ORIGIN  
 KIIIRKINGNNTNRVGGDILYFDMTINNKAYNLFMKNETWYADNHSTEDIYAICLR  
 EQTRDINDNIIFIQIPNNNTYYASQIFKSNFNGENISGICSGTGYRFLGGDWRHN  
 YLVPTVKQGNVALLSESTHMGFVPVSE"  
 QY 13 ACCATCCCATTCACATCTTCTCTACACCAACTCCCTGTTGAAGGACATCATCAAC 72  
 DB 2847 ACAATACCCCTTAATATTTTTCATATACTAATAATCTTTTAAAGATATATTAAT 2906  
 QY 73 GAGTACTTCAACAACATCAACGACTCCAGACTCCCTGTCGAGACCGGPAAGAACACC 132  
 DB 2907 GAATATTTCAATATATATATGATTCAAAAATTTTGAGCCCTACAAAACAGAAAAAT 2966  
 QY 133 TTGGTCGACACCTCCGGTTACAAACCGCAGGCTCCGAGGAGGCTGACGTCACGCTAAC 192  
 DB 2967 TTAGTGGATACATCAGATATATATGACAGACTGAGTGAAGGCGGATGTCAGCTTAAT 3026  
 QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGGTTCCTCCGGTGAGGACAGGPAAGGTCAATC 252  
 DB 3027 CCAATATTTCCATTTGACTTTAAATTAGGTAGTTTACAGGGGAGGATAGAGTAAAGTTATA 3086  
 QY 253 GTACCCAGACGAGACATCGTCTACAACTCCATGTACGAGTCTCTCCATCTCCATCTCCTC 312  
 DB 3087 GTACCCAGATGAATAATTTGTATATATTTCTATGATGAAAGTTTTCAGCATAGTTT 3146  
 QY 313 TGGATCAGAAATCAACAAGTGGGTCTCCAACTTGGCAGGTTACACCATCATCGACTCCGTC 372  
 DB 3147 TGGATTAGAAATTAATAATGGGTAAGTAATTTTACCTGGATATACTATAATGTAGTGT 3206  
 QY 373 AAGAACAACTCCGGTGGTCCCATCGGTATCATCTCCAACTTCTCTGGTCTTCAACCTCAAG 432  
 DB 3207 AAAATAACTCAGGTGGGATAGTATAGTATATTTAGTAAATTTTGTATTTTACTTTAAA 3266  
 QY 433 CAGACGAGGACTCCGAGCAGTCCCATCACTTCTCTACGACATCTCCAAACAGCTCCT 492  
 DB 3267 CAAATGAAGATAGTGAAACAAAGTATAAATTTTACTTATGATATCAAAATAGTCTCT 3326  
 QY 493 GGTTACAAAGTGGTCTTCTCGTCACCGTCACCAACACATGATGGGTAGCATCAAGATC 552  
 DB 3327 GGATACATAAATGGTTTTTTTCTAACTGTTACTAAATAATGATGGGAAATATGAAGAT 3386  
 QY 553 TACATCAACGGTAAAGTGTATCGACACCATCAAGGTCAAGGATTCACCGGTATCAACTTC 612  
 DB 3387 TATATAATGGAAATTAATAGATACTATAAAGTTAAAGAACTAACTGGAATTAATTT 3446  
 QY 613 TCCAGACCATCACTTCGAGATCAACAAGATCCAGACACCGGTCTGATCACTCCGAC 672  
 DB 3447 AGCAAACTATACATTTGAAATAAATAAATTCAGATACCGGTTTGTATTACTTACAT 3506  
 QY 673 TCCGCAACATCAACATGTGGATCCGTGACTTCTACATCTTTCGCCAAGGATTTGGACGGT 732  
 DB 3507 TCTGATAACATCAATGTGGATAGAGATTTTATATTTGCTTAAAGATAGATGGT 3566  
 QY 733 AAGGACATCAACATCTCTTCAACTCCCTTTCAGTACACCAAGCTCGTCAAGGACTACTGG 792  
 DB 3567 AAAGATTAATATATATTTTAAATAGCTTGCATATACTAATTTTGTAAAAGATTTATGG 3626  
 QY 793 GGTAAAGCTGAGATACAAAGGAGTACTACATGGTCAACATCGACTACTTGAACAGA 852  
 DB 3627 GGAATGATTTAAGATTAATAAAGAAATTAATATGTTTAAATAGATTAATTAATAAGA 3686  
 QY 853 TACATGTACGCCAACTCCAGACAGATCGCTTTCACACACGAGCTTAACAACAGACTTTC 912  
 DB 3687 TATATGTATGCGAACTCAGCAGAAAATTTGTTTTTAATACACGTAGAAAATAAATGACTTC 3746  
 QY 913 AAGCAGGTTACAGATCATCATCAGCGTATCAGAGGTATACACCAACGACACGAGCTC 972  
 DB 3747 AATGAAGGATATAAAATTAATAAATAAAGAAATCAGAGGAATACAAATGATAGATGA 3806

QY 973 AGAGTGTGACATCTCTGATCGATGACTATCAACAAGCGCTACACCTGTTTC 1032  
 DB 3807 CGAGGAGAGATATTTATATTTTGTATGACATTAATAACAAGCATATATTTGTTT 3866  
 QY 1033 ATGAAGACGAGACATCTAGCGGACAACTCCACCGAGGACATCTACGCCATCGGT 1092  
 DB 3867 ATGAAGATGAACATCTAGCGGACAACTCACTAGTACTCAAGATATATGCTATAGT 3926  
 QY 1093 CTGCTGAGACAGCAGGACATCAACGACAACTATCTTCAGATCCAGCCAAATGAAC 1152  
 DB 3927 TTAAGAGACAAACAAAGGATATAAATGATAATATATATTTCAATACAAACAAATGAAT 3986  
 QY 1153 AACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGACATCTCC 1212  
 DB 3987 ATACTTATTTAGCATCTCAATATTTAATCAATATTTAATGAGAAATATTTCT 4046  
 QY 1213 GATCTGTTCCATCGGACATCAAGATCTCGGTGCTGGTGTGATGACATCAACACAC 1272  
 DB 4047 GGAATATGTTCAATAGTACTTATCGTTTATAGCTTGGAGTGTGATGATAGACAAAT 4106  
 QY 1273 TACTTGTGTTCAACGTTCAACGAGGATCACTACGCTTCTGCTGGAGTCCACTTCCAC 1332  
 DB 4107 TATTTGTTGCTACTGTTGACGAGGAAATATGCTTCATATATAGATCAACATCACT 4166  
 QY 1333 CACTGGGATGTTCCCGAGTCTCCGAGTAA 1362  
 DB 4167 CATTTGGGTTTGTACCTGTAAGTGAATAA 4196

RESULT 3  
 CBPHCNCLM 4712 bp DNA linear BCT 23-DEC-1993  
 C.botulinum phage C mutant gene for neurotoxin type C1.  
 X71126  
 X71126.1 GI:414619  
 mutant; neurotoxin.  
 Clostridium botulinum.  
 Clostridium botulinum.  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.  
 1 (bases 1 to 4712)  
 Direct Submission  
 Hauser.D.  
 Submitted (03-MAR-1993) D. Hauser, Inst. Pasteur, Unite des Toxines  
 Microbiennes, 28, Rue du Docteur-Roux, F-75724 Paris Cedex 15,  
 FRANCE  
 2 (bases 1 to 4712)  
 Hauser.D., Gibert,M., Eklund,M.W., Boquet,P. and Popoff,M.R.  
 Comparative analysis of C3 and botulin neurotoxin genes and their  
 environment in Clostridium botulinum types C and D  
 J. Bacteriol. 175 (22), 7260-7268 (1993)  
 94042900  
 8226673

Location/Qualifiers  
 1. 4712  
 /organism="Clostridium botulinum"  
 /strain="type C (con toxic)"  
 /sub\_strain="A028-CN phage"  
 /db\_xref="taxon:1491"  
 791..796  
 807..4682  
 /gene="neurotoxin mutated gene"  
 807..4682  
 /gene="neurotoxin mutated gene"  
 /pseudo  
 /codon\_start=1  
 /transl\_table=11  
 /db\_xref="PID:e69305"  
 2463..2465  
 /gene="neurotoxin mutated gene"  
 /note="intrinsic stop-codon"  
 1907 a 482 c 721 g 1602 t

FEATURES  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 FEATURES  
 source

ORIGIN

Query Match 48.5%; Score 665.2; DB 1; Length 4712;  
 Best Local Similarity 68.3%; Pred. NO. 1.5e-92;  
 Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;  
 QY 13 ACCATCCCATTCACATCTCTCTCTCACCAACAACCTCCCTGTTGAAGGACATCATCAAC 72  
 DB 3333 ACAATACCCCTTATATTTTTCATATACATAAATCTTTTATTAAGAATATAAATAAT 3392  
 QY 73 GAGTACTTCAACACATCAACGACTCCAGATCTCTGCTCCCTGAGACAGCGTAGAACACC 132  
 DB 3393 GAATATTTCAATATATTAATGATCAAAAATTTTGGAGCTACAAACAGAAAAATCT 3452  
 QY 133 TTGCTGCACACCTCCGGTTACACGCCGAGGTCTCCGAGGAGGTGACGCTCCAGCTCAAC 192  
 DB 3453 TTAGTGGATACATCAGATATATATGACAGAGTGAAGTGAAGAGCGATGTTACGCTTAAT 3512  
 QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGGTTCCCTCGGTGAGGACAGAGTAGAGTCATC 252  
 DB 3513 CCAATATTTCCCATTTGACTTTAAATAGTAGTTCAGGGGAGGATAGAGTAAAGTTATA 3572  
 QY 253 GTCACCCAGAGAGGAAACATCGTCTCAACCTCCATGTACGAGTCTCTCCATCTCCTTTC 312  
 DB 3573 GTACCCAGATGAAT 3632  
 QY 313 TGGATCAGATCAACAAAGTGGTCTCCAACTGTCAGGTTACACCATCATCGACTCCGTC 372  
 DB 3633 TGGATTAGATATAAATAATGGTAAAGTAAATTTACCTGGATATATATATATATATATAT 3692  
 QY 373 AAGACAACTCCGGTTCCTCGGTATCATCTCCAACTCTCTGGTCTTCAACCTCAAG 432  
 DB 3693 AATAATACTCAGTTGGATAGTATATATATATATATATATATATATATATATATATAT 3752  
 QY 433 CAGAACGAGGACTCCGAGCAGTCCATCAACTTCTTACGACATCTCCAAACAGGCTCT 492  
 DB 3753 CAAATAGAGATAGTGAACAAAGTATAAATTTAGTTAGTATATATATATATATATATAT 3812  
 QY 493 GGTACACACAGTGTCTCTCGTCCAGTCCACCAACACATGATGGTACATCAAGATC 552  
 DB 3813 GATACAAAT 3872  
 QY 553 TACATCAACGAGTATGATCGACACCATCAAGTCAAGGATGACCGGTATCAACTTC 612  
 DB 3873 TATATAATATGAAAT 3932  
 QY 613 TCCAGACCATCACCTTCGAGATCAACAGATCCACACACCGGTCTGATCACTCCGAC 672  
 DB 3933 AGCAAACTATACATTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3992  
 QY 673 TCCGACACATCAACATCGGATCCGACTTCTACATCTTCCGCAAGGAGTGGAGGT 732  
 DB 3993 TCTGATAACATCAATATGTTGGATAGAGATTTTATATATATATATATATATATATAT 4052  
 QY 733 AAGGACATCAACATCTCTTCAACTCTTCCAGTACACCAAGCTCGTCAAGGACTACTGG 792  
 DB 4053 AAGAT 4112  
 QY 793 GGTAAACGACTCAGATCAACAAAGGAGTACTTACATGGTCAACATCGACTACTTCAACAGA 852  
 DB 4113 GGAAT 4172  
 QY 853 TACATGTAGGCAACTCCAGACAGATCGTCTTCAACACAGACCGTACACACAGCTTC 912  
 DB 4173 TATATGTATGCGAATCAGGACAAATTTTATATATATATATATATATATATATATATAT 4232  
 QY 913 AACGAGGTTCAAGATCATCATCAAGGTATCAGAGGTAAACCAACAGCAGACGATC 972  
 DB 4233 AATGAGGAT 4292  
 QY 973 AGAGTGTGACATCTCTGATCGATGACTATCAACAAGGAGGCTTACACCTGTTTC 1032  
 DB 4293 CGAGGAGGAGATATTTTAT 4352



QY 1033 ATGAAGACGAGACCATGTACGCCGACACCACTCCACCGAGGACATCTACGCCATCGGT 1092  
 DDb 4353 ATGAAGATGAACATATGATGACAGATATCATAGTACTGAAGATATATATGCTATAGGT 4412  
 QY 1093 CTGCGTGAGCAGACCAAGACGATCAACGACACATCATCTTCCAGATCCAGCCCAATGAAC 1152  
 DDb 4413 TTAAGAGAACAAACAAAGGATATAAATGATATATATATTTCAATACACCAATGAAT 4472  
 QY 1153 AACACTTACTACTAGCTTCCCGAGATCTTCAAGTCCAACTTCAAGGCTGAGAACATCTCC 1212  
 DDb 4473 AATACCTATATATAGCATCTCAATATATTTAAATCAATTTTATGAGAGAAATATTTCT 4532  
 QY 1213 GGTATCTTCCATCGGTACCTACAGATTCGGTCTGCGTGGTGAGTGTGACGACACCAAC 1272  
 DDb 4533 GGAATATGTTCAATAGGTACTTATGCTTTAGACTTGGAGGTGATGCTATAGACACAAAT 4592  
 QY 1273 TACTTGGTCCAACTGCTCAAGCAGGTACACGCTCCCTCTGCTGGAGTGCACATCCACC 1332  
 DDb 4593 TATTGGTGCCTACTGTGAGCAAGAAATATGCTTCAATATTAGAAATCAACATCAACT 4652  
 QY 1333 CACTGGGATTCGTCOCAGTCTCCGAGTAA 1362  
 DDb 4653 CATTTGGGTTTGTACCTGTAAGTGAATAA 4682

RESULT 4  
 CBCTOX  
 FOCUS  
 DEFINITION Botulinum bacteriophage genes for HA-17, HA-33, nontoxic components and C1 toxin.  
 ACCESSION X62389  
 VERSION 1  
 KEYWORDS botulinum toxin; C1 toxin; HA-17 gene; HA-33 gene; Hemagglutinin; neurotoxin; nontoxic; toxin.  
 SOURCE Clostridium botulinum.  
 ORGANISM Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.  
 REFERENCE 1 (bases 1 to 9613) Tsuzuki, K., Murakami, T., Indoh, T., Kimura, K., Fujii, N., Yokosawa, N., Syuto, B. and Oguma, K.  
 TITLE The complete nucleotide sequence of the gene coding for botulinum type C1 toxin in the C-ST phage genome  
 JOURNAL Biochem. Biophys. Res. Commun. 171 (3), 1304-1311 (1990)  
 MEDLINE 91024998  
 PUBMED 2223445  
 REFERENCE 2 (bases 1 to 9613) Tsuzuki, K., Kimura, K., Fujii, N., Yokosawa, N. and Oguma, K.  
 AUTHORS Tsuzuki, K., Kimura, K., Fujii, N., Yokosawa, N. and Oguma, K.  
 TITLE Nucleotide sequence of the gene for one of the components of hemagglutinin produced by Clostridium botulinum type C  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 9613) Tsuzuki, K.  
 AUTHORS Tsuzuki, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-DEC-1991) K. Tsuzuki, Sapporo Medical College, Dept of Microbiology, South 1 West 17, Sapporo 060, JAPAN  
 COMMENT On Oct 15, 1994 this sequence version replaced gi:40386.  
 See also X53041 and D90210.

FEATURES  
 Location/Qualifiers  
 1. 9613  
 /organism="Clostridium botulinum"  
 /strain="type C Stockholm, c-st"  
 /db\_xref="taxon:1491"  
 complement(111..551)  
 /gene="HA-17"  
 complement(111..551)  
 /gene="HA-17"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA44260.1"  
 /db\_xref="GI:40387"  
 /db\_xref="SWISS-PROT:P46083"

/translation="MSSERTFLPNQYKIKSLFSDSLTYTSSGALSFNTSLDNQK  
 WKLYISSSNGFRFSNVAEPNKYLAVNDYGFYILSSSSNNSLWNPTKIAINSYICLT  
 SVINVTYANTYIDNNNNITDQPTLNLNPNEDINNNOILKLEKL"  
 complement(560..565)  
 complement(614..1474)  
 /gene="HA-33"  
 complement(614..1474)  
 /gene="HA-33"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA44261.1"  
 /db\_xref="GI:40388"  
 /db\_xref="SWISS-PROT:P46084"  
 /translation="MSOTNANDLNKNEVEFFISPSNNVKNVLDKISOSEVKLWKLKSLG  
 NQKRLYIDTNKQAYIKVMDNTSLILITWNPASSVSVDKTDNGDQYWTLLQNYLSR  
 NVIRNTNPNLVLQYINDTLMVSTSSNQFFXFSNCIYEAALNNRCKLQTLNS  
 DRFLSKNLQIIVLWQFDSRQKWIIEYNETKSAYTLKQENNYLWQINSNYV  
 EYQSTSLQIYWNINLNDNASKYILYLNQDTRVLDVYNSQIANGTHVIVDSYHGN  
 TNQWMIINL"  
 complement(1482..1486)  
 1726..1731  
 1737..5327  
 /gene="nontoxic"  
 1737..5327  
 /gene="nontoxic"  
 /codon\_start=1  
 /transl\_table=11  
 /product="non-toxic components"  
 /protein\_id="CAA44262.1"  
 /db\_xref="GI:40389"  
 /db\_xref="SWISS-PROT:P46081"  
 /translation="MDINDLNINSVDNKNVIVRARKINTFFKAFKVPAPNIWVAPE  
 RYGEPLDAEYKLDGDIYDSNLFSDSERENFLOAIILLKRINNTISGKOLLILI  
 STAPFPGYIGGYSNPFTGKTPKSNKLSLNTSTIPFGGYRETNYESON  
 KNFTASNIIIFGGSNIVENVYIYKKAENAGCMGMAEIVQPLITKYNKYIDP  
 AMELTKLIKSLYGLIKPSDNLVYPLRLTDLNKFQSLNIDLLISGGVDELF  
 NLNFTQSFNSIIPDFSNALKHFKYQYTYMDYTNINGNFVNGQINTKPLSNL  
 TNIASKKVLNVLNENNISLMKSNYDGLKTTEDFYTKIPYNEVEYFNFSD  
 NEPLNNTISEVDSIPBIDINPKDNDNLVFTQITSMTEVTTLTINYLQAI  
 TNNEFTLSDFSKVSQKSLVYFGLNLMVLETIKNDGPIDTDKYLILKGVF  
 KNYFDNLQGEIDSMGINSYVLFGLKALNLTNSFVEYQDSGLYSLSKDL  
 REPIEDIDISLGSFQDLNKKLYEISKNIVYFKKIYFSLQWTEYISOYE  
 LICMAKSLAQESICVFEDYIPKFSIYMEYIINNINIKREFIORTNIDNSQSIS  
 NRYDNLNKAISICVFEDYIPKFSIYMEYIINNINIKREFIORTNIDNSQSIS  
 NSYFTKIDFQIDQIKNFQSVQVMEKILSPYLLFLFASKPNSNIIDISGK  
 NTLQYTESIELYGVNGESLYLAKSNETIKFSNKFTGLTNFTICFWLRFTGND  
 DKRLIGNVNCWEIYFEDNGLVFRIIDSGNOSVLSYLSNIINDNYYISISVDR  
 KDQLIFINDKNAVNSIDOLISYLSNLSYLNKNSIYVEELVDNPIITSEVIR  
 NYFDLNSYRDSKSLLEKNYQILNYVFPETSLYENDNKNKSYLSLKNFDGNI  
 SSVFKLINDESKYVOKWDECTICVDGTGKTLDISPENNRIQLVSKDNKAKTIV  
 NTDLFRPDCITFSYNDKYFSLSLRGGDYNMNCNDNNKVPKGAHLWILES"  
 5330..5335  
 5342..9217  
 /gene="C1 toxin"  
 5342..9217  
 /gene="C1 toxin"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA44263.1"  
 /db\_xref="GI:40390"  
 /db\_xref="SWISS-PROT:P18640"  
 /translation="MPITNNFYSDVDNKNLILDTLNTLANSEKAPRTTGNW  
 VIPDRFSNPNLKNKPRVTSKASYDPNYLSTDSKDTFLKEILKPKRINSEI  
 GEELIYRLSTDIPFGNNNTPTDFDVFNSVDKTRQGNWKNKTSGLNPVLTIG  
 PRENIDPETSTFKTLNNTFAAQGFALSIISIPRMTLYGNATNDVGEFRKSE  
 FCMPIILIMHELMNAHLNLYGAIIPNDQITSSVTNIFTYSQYNKLEIYAFGGP  
 TIDLIPKARKYFEKALDYRSIAKRLNSTTANPSSFNKYGKQKILIRYRVV  
 ESSGEVTNRNKFVELYNETQIETFEYNAKYNQNKIYLSNVTPTANTLDNV  
 YDIQGNFNRKSLNLYLPMGNLSRNPALRKAPENKLYLFTKCHKADGRSLYKNT  
 LDRELIVKTDLDFIGDISDKTDIFLRKDINEETEVYIPDNVSDVQVILSKNTSE  
 HGOLDLYPSIDSESLPGENQVYDNRQNDVYLSNYLESQKSLDNVEDFTTR  
 STEALDNSAKVYTYFTFLANKNAGVQGGFLMWANDVVEDEFTTILRLDKLIDK

VSALIPYIGPALNISVRGNFTFAVGTIILEAFPEFTIPALGAFVIYSKQVE  
 RNELITDNCLEORIKRNDSDVEMWGMWLSRIITOPNNTSYOMVDSINYQAGAIKA  
 KIDLEYKSSGSKENIKSOVENLNSLDVKISBAMNINIFIRCSTYIEFKWMLPK  
 VIDELNEEDNTAKLILIDSHNILLVGEVDKLRKAKVNSPONTIPNIFSYIYNNLS  
 LKDIINEYFNINDKILSLQKRNKTLVDTSYNAEVEEQQVLPNIPFPDFKLKSS  
 GEDRGKIVYVONEDSVNMFYSFSTFSPNIRKWSNLPVYIIDSXVNSGWSIGI  
 ISNPLVETLONEDSQSINFSVDISNAPGVNKFVTVNMMNKIYVINGKLID  
 TIKVKELTGINFSKTITFEINKIPDTGLITSDNINMWIRDFYIFAKELDKDNLIL  
 FNSLOVTVVYKDYNGNDLYNKEYYVYNDILYLNRYMANSRQIYVNTRENNDFNEGY  
 KILIKIRGNINTRVRGDDILYFDMTINKKAYNLFEMNEIMYANSHSTEDIYALGR  
 EYKADINLIIFQPMNNNTIYASQIFASNFNGENISGICSGIRFIRFLGDMTIRHN  
 YLPTVKQGNYSLLESTSHWGFVPSVSE"  
 BASE COUNT 3846 a 1003 c 1348 g 3416 t  
 ORIGIN

Query Match 48.5% Score 665.2; DB 1; Length 9613;  
 Best Local Similarity 68.3%; Pred. No. 1.4e-32;  
 Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;  
 13 ACCATCCCATCAACATCTTCCCTACACCAACATCCCTGTTGAAGGACATCATCAAC 72  
 7868 ACAATACCCCTTAATATTTTTCATATACATAAATCTTTATTAAGAGATATAATTAAT 7927  
 73 GAGTACTTCAACACATCAACGATCCAGATCTGCTCCCTGCAGAACCTTAAGACACC 132  
 7928 GAATATTTCAATATTAATGATTCAAAATTTTGAGCCTCAACAAACAGAAAATACT 7987  
 133 TTGGTCGACACCTCCGGTTACACGCCGAGGCTCCGAGGAGGTTGACGCCAGCTGAAC 192  
 7988 TTATGGGATACATCAGGATATAATGCAAGAGTGTGAGTGAAGAGGATGTTTCAGCTTAAT 8047  
 193 CCAACTTCCCATCCAGCTCAAGCTGGTCTCCCGTGCAGGACAGAGGTAAAGTCATC 252  
 8048 CCAATATTTCCATTTGACTTTAAATAGGTAGTTTCAGGGAGGATAGAGTAAAGTTATA 8107  
 253 GTACCCAGAGAGGAGACATCGTTACAACTCCATGACGAGTCCCTTCTCCATCTCCTTC 312  
 8108 GTAACCCAGATGAATATTTGATATATAATTTCTATGATGAAGCTTTTAGCATTTAGTTT 8167  
 313 TGATCAGATCAACAGTGGTCTCCAACTTGCAGGTTACACCATCATCGACTCCGTC 372  
 8168 TGATAGATGAATGAATGAGTGAATTTACTCGATATACATATAATGATGATGTT 8227  
 373 AAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTCCCTGCTTCCACCCGGAAG 432  
 8228 AAAATAACTCAGGTGGAGTATAGTTATTTAGTAAATTTTAGTATTTACTTTAAAA 8287  
 433 CAGAACGAGGACTCCGACGATCCATCACTTCTCCACGACATCTCCACACAGCTCCT 492  
 8288 CAAATGAAGATAGTGAACAAAGTATAAATTTTAGTTATGATATATCAATAATGCTCCT 8347  
 493 GGTACAAAGTGGTCTTCCGTCACCGTCAACCAACATGATGGTGAACATGAAGATC 552  
 8348 GGATACATAAATGCTTTTGTACTTACTTAAATATGATGGGAATATGAAGATT 8407  
 553 TACATCAAGGGTAACTGATGACACCATCAAGTCAAGGAGTTGACCGGTATCACTTC 612  
 8408 TATATAATGGAAATTAATAGATCTATAAAGTTAAAGTAAAGTAACTGGAATTAATTT 8467  
 613 TCAAGACCATCACCTCGAGATCAACAGATCCGACACCGCTCTGATCACTCCGAC 672  
 8468 AGCAAACTTAACTTTGAATAAATAAATCCAGATACCGGTTGATTTACTTCCAGAT 8527  
 673 TCGGACACATCAACATGTTGGATCGGTGACTTCTACATCTTCGCCAAGAGTTGGACGGT 732  
 8528 TCTGATAACATCAATATGTTGAAGATGATTTTATATATTTGCTAAAGATTTAGATGGT 8587  
 733 AAGGACATCAACTCGTGTCAACTCCTTGCAGTACACCAACGCTCGTCAAGGACTACTGG 792  
 8588 AAGATATTAATATATTTTAACTGCTGCAATATCTATATTTGTTAAAGATTTATGG 8647  
 793 GGTAAAGCCTGAGATACAAACAGGAGTACATATGTTGAACATCTGACTTCTGAAACAGA 852

Db 8648 GGAATGATTAAAGATATAATAAAGAAATATATATGTTAATATAGATTATTAAATAGA 8707  
 Qy 853 TACATGTACGCCAACCTCCACAGATCGTCTTCAACACACAGAGGTAAACACACGACTTC 912  
 Db 8708 TATATGTGCGAACTCACCAGCAAAATGTTTTTATACACGTAGAAATATATATGACTTC 8767  
 Qy 913 AACGAGGGTTACAGATCATCATCAAGCGGTATCAGAGGTAAACACCAACGACACGAGTC 972  
 Db 8768 AATCAAGGATATAAATATATAAATAAAGAAATCAGAGGAAATACAAATCATATAGATA 8827  
 Qy 973 AGAGTGTGACATCTCTGACTTCTCGACATGACTATCAACACACAGGCTTACACCTGTTTC 1032  
 Db 8828 CGAGGAGGATATTTTATATTTTGTATGACAAATTAATAACAAAGCATATATTTGTTT 8887  
 Qy 1033 ATGAAGAACGAGACCATGTACGCGGACCAACCTCCACCGAGGACATCTACGCCATCGGT 1092  
 Db 8888 ATGAAGATGAACATATGTATGCAGATAATCATAGTACTGAAGATATATATATAGGT 8947  
 Qy 1093 CTGGTGAGCAGCAGGACATCAACGACCAACATCATCTTCCAGATCCAGCAATGAAC 1152  
 Db 8948 TTAAGAACAAACAAAGGATATAAATGATATATATATTTTCAATACACCAATGAAT 9007  
 Qy 1153 AACACTTACTACTACGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGACATCTCC 1212  
 Db 9008 AATACTTATTTAGCATCTCAAAATATTTAAATCAAAATTTTAAATGAGAAAATATTTCT 9067  
 Qy 1213 GGTATCTGTTCCATCGGTACTACAGATTCGCTGCTGGTGGTGGTACTGGTACAGACACAC 1272  
 Db 9068 GGAATATGTTCAATAGTACTTATCGTTTGTAGACTTGGAGGTGATGGTATAGACAAAT 9127  
 Qy 1273 TACTTGGTTTCCAACTGTCAAGCAGGTTAACTACGCTCTCTGCTGGAGTCCACTTCCAC 1332  
 Db 9128 TATTTGGTGCCTACTGTGAAGCAAGGAAATATGCTTCATTATTAGTAATCAACATCAACT 9187  
 Qy 1333 CACTGGGATTCGTCGCCAGTCTCCGAGTAA 1362  
 Db 9188 CATTTGGGTTTGTACCTGTAAAGTGAATA 9217

RESULT 5  
 CBP1CNTC1 Clostridium botulinum phage 1C, CHN-14, CHN-33, CHN-138 and BONT/C1  
 LOCUS  
 DEFINITION  
 ACCESSION X65433.1 GI:509271  
 VERSION X65433.1  
 KEYWORDS haemagglutinin; neurotoxin.  
 SOURCE Clostridium botulinum.  
 ORGANISM Clostridium botulinum  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.  
 REFERENCE 1 (bases 1 to 9689)  
 AUTHORS Hauser, D.F., Eklund, M.W. and Popoff, M.R.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 9689)  
 AUTHORS Hauser, D.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-MAY-1992) D.F. Hauser, Institut Pasteur, Unite des  
 Toxines Microbiennes, Institut Pasteur, 28 rue du Docteur Roux,  
 F-75724 Paris Cedex, FRANCE  
 FEATURES  
 source  
 1. 9689  
 /organism="Clostridium botulinum"  
 /db\_xref="taxon:1491"  
 complement(253..647)  
 /gene="CHN-14"  
 complement(253..633)  
 /gene="CHN-14"  
 /codon\_start=1  
 /transl\_table=11  
 /product="14kDa protein associated with BONT"  
 /c1-haemagglutinin complex"  
 /protein\_id="CA47057.1"

```

/db_xref="GI:509272"
/db_xref="SPRMBL:Q45915"
/translation="MSERTEPFGNKKKSLFSDSLYLYSSGALSFSNTSSLDNOK
WKLEVISSNGFESNAEENKVLAYNDYGVILSSSSNSLWNPKIKAINSYICTL
SIVNVTYATWIDNNNNIHDLP"
complement(642..647)
/gene="Chn-14"
/gene="Chn-15"
/gene="Chn-33"
/gene="Chn-33"
/codon_start=1
/trans_table=1
/product="33kDa protein associated with Bont
/C1-haemagglutinin complex"
/protein_id="CAA47058.1"
/db_xref="GI:509273"
/translation="MSTANDLRNNEVFIPSSNNTKVLKDISOEVKLNKLSGA
NOKRLVLDNKAQYKIKVMDNPSLITNNAFLSSVSVKTDNGNOYVLLQNYISR
NVIIRTNMNNLVLOYNIDTTLAVSTOTSSNOFFPSCNIEALNRRCKLOTOLNS
DRFUSKLNQSIIVLWQWFSRQKIIETNYSATYLLKQENNRNRLYVQNSNYV
ETYSQSLIQYWNLYLDNDASKYILYLNQDTNRVLDVYNSQIANGTHVIVDSYHGN
TNOQWIIINLI"
complement(1564..1568)
/gene="Chn-33"
/gene="Chn-33"
/gene="Chn-33"
/gene="Chn-33"
/join(1803..1808,1819..5409)
/gene="Chn-138"
/gene="Chn-138"
/gene="Chn-138"
/gene="Chn-138"
/codon_start=1
/trans_table=1
/product="138kDa protein associated with Bont
/C1-haemagglutinin complex"
/protein_id="CAA47059.1"
/db_xref="GI:4376225"
/translation="MDINDLNINSPVDNKNVIVRARKNTNFFKAFVAPNITWAVE
HYVEGPDIAEEYKLDGIYDSNPLSDSRENFLOAILILKRINTTISGQLLSLI
STALFFPIGIGGSFNFTFGKPKSKKJNSLVTSIPEFPGYREINIBSON
KNKVAIVIGGSSNENNVYIYKKNDNAEMGTMAEIVFPQLTYKYNKFIIDP
AMELTKLISLYELGKIPSDNLVPEYRLTELDNKPQSLNIDILLISGVDLEFI
NTNPTWNTSFPNSIKMFYKNIYKTEEGNAIGNDIKRLKQKFOINQVDIWNL
NLNFKQSFNSIIPDRSNALKHFYKQYTYMDYTNINGVNGQINIKPLSNKN
TNISKPEKVNLVNENNISLMSKNYIGDGLKFTEDFYSTYKIPYNEEYEFNDSD
NPLNNISLEVDSPILIDINPKNSDNLVETQISMEETHTALSINYLQAI
TNENFPLISDFSVWSKSKSLVYSELDNMLSYLETIKNDGFDIDKKIYLWLKEV
ANYSPDNLTOEIDSMGNEVLFWSKALNLTNSFVEEYQDSGALSLSKKDNL
REPIEDIDSDLSLGSFGLDNKKLYEISKNVYFKIYFSLDQWTEYISQYE
LICMAKQSLAQESLVQVQNKPTDLSKASIPDTLKIRETFKTFIDLNESQIS
MNRVDNFLNKASICVQVEDIYKPIFISMEKYINNINIKYREFTORNTINDNESKILI
NSYTFKTIIDRFIDQISIKNFNSOVQVMEILSPVOLLFPASKGPNILIEDISGK
NTLLQYVESIELYGVNGESLYLKSNETIKFSNKFTGLNNTFTICFVLSFTKND
DKTLILGNKNKNCGWEIYFEDNGLVFELIISNGNSVYLSINDNWNYYIISVDRL
KQDLIFINDKNVNSIDQLLSYSTNIIISLVNKNNSIYVEELSVDNPITSEEVIR
NYFSLDYSIIRDSKSLNKNQYNYVFPETSLYEVDNKNKSLYLSKNTDGINI
SSVFKLINIDESYVQKWDCEICVLDGTEKYLIDSPENNRIOQLVSSKDNKAKTIV
NTDLFRPDCTIFSYNDKYFSLSRDGNVNMICNDNKNVPKGAHILLES"
5408..5413
/gene="Bont/C1"
5408..5413
/gene="Bont/C1"
5424..9299
/gene="Bont/C1"
5424..9299
/gene="Bont/C1"
/codon_start=1
/trans_table=1
/product="Botulinal C1 neurotoxin"
/protein_id="CAA47060.1"
/db_xref="GI:509275"

```

```

/db_xref="SWISS-PROT:P18640"
/translation="MPITINNFNSYSDVDNKNILYLDTHLNIANEPEKAFRITGNIN
VIDPFRNPNLNKPPRVTPSKGYDPNTSLTSDSDKDTLKEIKLFRKINSREI
GEELVRLSDIIPGNNNTPIINTPDDVDNSVDKTRGNNWVKTGINSFVITIG
PRENIDPSTFKLPNNTPAQEGGALSILISPRFMLTYSNATNDGGRFQSKSE
FCMDPILILMHLNHLNHLNLYGAIENDQTISSVTSNIYSQYNKLEAEIAYGGP
TIDIPKSAKYFEERKALDYIRIARLNSITITANPSSFNKYGFKQLIRKIRFVV
ESSEGVNARKNFVELYNELTOIFTEFNKYNQNRKYLNSVTPVTANILDNV
YDQNGENIPKSNLNLFMQNLSPALRKVNENMLYFTKFKHAKIDGRSLYKNT
LDGRELKNTDLPFGIDISDKTIDFLRKDINEETEVIYYPDNVSQDVLISKNTSE
HGOLDLYPSIDSESELPGENOVFYDNRQNVYLSNVYLSQSLSDQVEDFTFR
SIEBALNSAKVYTYPTLANKVAGVQGLFLMWANDVDEFTNLRKDLKIDSD
VSAIIPYIPALNAINSVARGNEFAVGTGVTILLAEPESTIPALGAFVYLSKVOE
RNEIKITIDCLQRIKRWKSDSEMMGWLSRIITQNNISQYMCUSLNYAGAIKA
KIDLEYKYSKDKENIKSQVENLSDVKISEAMNNLNKFRSEYSLYFNKMLPK
VIDELNEFDTRAKLINILDSHNIILYGEVDKLRKAVNNNSFQNTIPFNIFTNNSL
LKDINEFNNINDSKILSLQNKNTLVDTSYNAEVSSEGVQNLNPFPPFKLGSS
GEDGRKVIYTONENIVNYSYBSFISNAPRYNKVSNLPGYTIIDSVKNNSGWSIGI
ISNPLVTLAKONDSQSINFSYDISNAPRYNKVSNLPGYTIIDSVKNNSGWSIGI
TIKVKELTG.NFSKTIIFELINKIPDGLITSDSDNINMLRDEYIFAEKELDKDNLIL
FNSQIYTNVNDWGLNDLRNKYIYVNDIYLRNMYANSRQIVYTRNNRDNFEGY
KILIRGRNTDTRVGGDILYDMTINNNKATLPMKNETMADNHSDEDIYALGLR
EQTKDINDNTIFQIQPMNNTYYASQIFKSNFNGENISGICISIGTYRFLSGDWYRHN
YLVPYVQGNVYALLSESTHWGFVPVSE"
9385..9424
/gene="Bont/C1"
9385..9424
/gene="Bont/C1"
/feature="putative"
BASE COUNT 3865 a 1022 c 1359 g 3443 t
ORIGIN
Query Match 48.5%; Score 665.2; DB 1; Length 9689;
Best Local Similarity 68.3%; Pred. No. 1.4e-92;
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;
QY 13 ACCATCCCATCAACATCTTCTCCTACACCAACACCTCCCTGTGAGAGGACATCAAC 72
Db 7950 ACATACCCCTTAATATTTTTCATATACTAATAATCTTTATTAAAGATATAATTAAT 8009
QY 73 GAGTACTTCACACATCAACGACTCCAGATCTCTCCCTGCAGAACCGTAAGAACACC 132
Db 8010 GAATATTTTCAATATTAATTAATGATTCAAAATTTTGACCCCAACAAAGAAAATACT 8069
QY 133 TTGGTCGACACCTCCGGTTTCAACGCGAGGTCTCCGAGGAGGTGACGTCCACCTCAAC 192
Db 8070 TTAGTGGATACATCAGGATATAATGACAGAGTGAAGGAGCGATGTTCAGCTTAAT 8129
QY 193 CCAATCTTCCCATTCGACTTCAAGTGGGTCTCCCGTGGAGGACAGGTAGGTGATC 252
Db 8130 CCAATATTTCCATTTGACTTTAAATAGGTAGTTCAGGGGAGGATAGAGGTAAAGTTATA 8189
QY 253 GTCCACCAAGACGAGAACATCGCTTACAACTCCATGTACGAGTCTCTCTCCATCTCTTC 312
Db 8190 GTAACCCAGATGAATAATATGTATATAATCTATGATGAAGCTTTTAGCATTTAGTTT 8249
QY 313 TGGATCAGATCAACAGTGGGTCTCCAACTTCCAGGTTCACACCATCAGCTCCGCTC 372
Db 8250 TGGATTAGAATAATAATAATGGGTAGTAAATTTTACCTGGATATACATAATGATAGTGT 8309
QY 373 AAGAACACTCCGGTTGTTCATCGGTATCATCTCCAACTTCCTGGTCTTCCACCTCAAG 432
Db 8310 AAAATACTCAGTGGAGTATAGGTATATTAGTAATTTTATTAGTATTACTTTTAAA 8369
QY 433 CAGACGAGGACTCCGAGCAGTCCATCAACTTCTCCTACGACATCTCCACACAGCTCCT 492
Db 8370 CAAATAGAGATAGTGAACAAAGTATAAATTTTATGTTATGATATATCAATAATGCTCCT 8429
QY 493 GGTTCACACAGTGGTCTTCGTCCCGCTCCACCAACACATGATGGGTGATCAAGATC 552
Db 8430 GGATACATAAATAGTGTGTGTGTAAGTCTTACTACATAATGATGGGAATATGAAAGATT 8489
QY 553 TACATCAACGGTAAAGTGTATGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTC 612

```

Db	8490	TATATAATGAAAATTAATAGACTATATAAAGTTAAAGAACTAACTGGAATTAATTTT	8549
QY	613	TCCAGACCATCACCTTCAGATCAACAAGATCCAGACACCGGTCTGATCACTCCGAC	672
Db	8550	AGCAAACTATACATTTGAATTAATAAATCCAGATACCGGTGATCTTCAGAT	8609
QY	673	TCCGACAACTACATCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGAT	732
Db	8610	CTGTGATACATCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	8669
QY	733	RAGGACATCAACATCTCTTCAACTCTTCAAGTACACCAAGCTGCGAAGGACTACTG	792
Db	8670	AAAGAT	8729
QY	793	GCTAACGACCTGAGATACACCAAGGATGATGATGATGATGATGATGATGATGATGATGAT	852
Db	8730	GGAAATGATTTAAAGATATAAAGAAATATATATGTTGATATAGATTTAAATAGA	8789
8	853	TACATGTAGCGCACTCCAGACAGATGCTTCAACACGAGCTGACCAACGACTTC	912
8	8790	TATATGTAGCACTACGACAAATTTGTTTAAATACAGTAGAATAATATATGACTTC	8949
8	913	AACGAGGTTTCAAGATCATCATCAAGCGTATCAGAGGTAAACACGACGACACGAGTC	972
8	8850	AATGAAGGATATAAAATTAATAAAGAAATCAGAGGAAATACAAATGATAGTAGTA	8909
8	973	AGAGGTGTGATCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	1032
8	8910	CGAGGAGGAGATATTTATTTGATATGACAAATTAATAACAAAGCATATATTTGTTT	9869
8	1033	ATGAAGAGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1092
8	8970	ATGAAGATGAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	9029
8	1093	CTGCGTGACCAACCAAGGATCAACACCAACATCATCTTCCAGATCCAGCCCAATGAAC	1152
8	9030	TAAAGAGAACCAAGGATATAAATGATATATATATATTTCAAAACACCAATGAAT	9089
8	1153	AACACTTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGCTGAGACATCTCC	1212
8	9090	AATACTTATTTACGATCTCAATATTTAAATCAAAATTTAATGAGAAATATTTCT	9149
8	1213	GGTATCTGTTCCATCGGTACGATGATGATGATGATGATGATGATGATGATGATGATGAT	1272
8	9150	GGAATATGTTCAATAGGATCTATGTTTATGATGATGATGATGATGATGATGATGATGAT	9209
8	1273	TACTTGTTCCTCACTGTCAGGAGGATCACTACGCTCTTCTGCTGAGTCCACTCCACC	1332
8	9210	TATTTGTCCTTACTGTGAGCAAGAAATTTATGCTTATTTAGATCAACATCAACT	9269
8	1333	CACGTGGGATTCGTCCTGATCTCCAGTAA	1362
8	9270	CATTGGGGTTTTGTACCTGTAAGTGAATAA	9299
RESULT	6		
AB061780			
LOCUS	AB061780	11747 bp	DNA linear BCT 06-NOV-2001
DEFINITION	Clostridium botulinum orf-22, ha-70, ha-17, ha-33, ntaha, nt genes for ORF-22, HA-70, HA-17, HA-33, NTNHA, neurotoxin, complete cds.		
ACCESSION	AB061780		
VERSION	AB061780		
KEYWORDS	AB061780.1	GI:16580753	
SOURCE	Clostridium botulinum (strain:C-Voichi) DNA.		
ORGANISM	Clostridium botulinum		
REFERENCE	Sagane, Y., Kouguuchi, H., Watanabe, T., Sunagawa, H., Inoue, K., Fujinaga, Y., Oama, K. and Ohshima, T.		
AUTHORS	Role of C-terminal region of HA-33 component of botulinum toxin in hemagglutination		
TITLE			

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

Biochem. Biophys. Res. Commun. 288 (3), 650-657 (2001)  
21534265  
2 (bases 1 to 11747)  
Sagane, Y., Watanabe, T., Kouguuchi, H., Morita, M., Miyamoto, T., Yamamoto, T., Muroga, A. and Ohshima, T.  
Direct Submission  
Submitted (16-MAY-2001) Toshihiro Watanabe, Tokyo University of Agriculture, Faculty of Bioindustry, Yasaka 196, Abashiri, Hokkaido 099-2493, Japan (E-mail: t-watanabe@bioindustry.nodai.ac.jp, Tel: 81-152-48-3843, Fax: 81-152-48-3843)  
Location/Qualifiers  
1. 11747  
/organism="Clostridium botulinum"  
/strain="C-Voichi"  
/db\_xref="taxon:1491"  
6. 545  
/gene="orf-22"  
/gene="orf-22"  
/codon\_start=1  
/transl\_table=11  
/product="ORF-22"  
/protein\_id="BAB71744.1"  
/db\_xref="GI:16580754"  
/translation="MNDLFYAEINLKHNDNQNFNFIEMSLKKYIEKTKKYNLYDYNN  
DILYHLMKELIENLNFSELDLRKYSTIKRYCINICKKNDKIIYNSEVYK  
KLDVAVVSLYCNFELDLISILNYKEKQIIYMKFEFGKDNELAIRLSRSIYK  
IRKSLKLYPIVMQLVNI"  
/complement(620..2491)  
/gene="ha-70"  
/complement(620..2491)  
/gene="ha-70"  
/codon\_start=1  
/transl\_table=11  
/product="HA-70"  
/protein\_id="BAB71745.1"  
/db\_xref="GI:16580755"  
/translation="MSLSIKELYYTKDSINNVNLAGVNVNRGDNILSRQNLG  
GNISNGCTAIVGLRIRETATPYIPTAFSEEEIKNNVQVNFANFEASIPGFE  
FSKTPAKSLMYLQTYIRYIEIKVLQNTVERAVLYVPSLVYKSEFNSSEQID  
KNFTYQSDKILNEKFIYKIDDTIVKESNNNNINFTSQILPNSLYVINK  
GDGVMNDKDLIGLIEESSTGSIQPLRNTTLPENTSNPTISQETEARLND  
AFNLQNTSTTLFKFVEAPTKNKSVINTYKVELYNQNGINDKABYLPLSL  
GKCVSPAPSPQVAVETPDQGTQTPNENIIVGVINSENIEESTIPDDDTY  
NIPISQNNACVLFKVNVTGVYKTKNNLPLIIEAIGSSNNMNSNNIK  
AIRITGKNSDKASLIVSLFKKNYIRIPIQSSSTSQLFKRELGNISLADST  
VNLNLTSGTHYTRQSPDNGNYSIQITPDGFNNFIASSIFSPRTNNNGIGTLY  
RLTESINGYNIITINNYSDLLNVEPISLNGATYIFRVKVTLENNYNIIFDAIRNS"  
/complement(2505..2945)  
/gene="ha-17"  
/complement(2505..2945)  
/gene="ha-17"  
/codon\_start=1  
/transl\_table=11  
/product="HA-17"  
/protein\_id="BAB71746.1"  
/db\_xref="GI:16580756"  
/translation="MSSRTFLPNGYKIKSLFNSLYLTYSAGALSFGNTSLDNQK  
WKLEIVSSNGFSPNVAEPKNVLAIVNDYGVIVLSSSSNNSLWNPKIANSYIICLT  
SIVNVTYATITVDNNNITDQILNLPNFIDNNNQILKLEKL"  
/complement(2999..3856)  
/gene="ha-33"  
/complement(2999..3856)  
/gene="ha-33"  
/codon\_start=1  
/transl\_table=11  
/product="HA-33"  
/protein\_id="BAB71747.1"  
/db\_xref="GI:16580757"

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS



gene ANIP-10

[illegible]

```

MNRVDNFKASICVFVEDIYPRKFIYSMEKYINNINIKITREFIQRCTNINDNEKILII
NSYFKIDRKFIDIOSIKNFNSOEVOMKEILSPVOLLFASKGPNINILEDISGK
NTLQVYESIELVANGESLYLKSPNERIKFSNKFTNGLNFTICPWLFTSVDND
DKTLGKNNKNGWVEYFEDGLYFEIIDSNGNQESVYLSHILINDNWIYFISGKND
KQLLIFINDNANWSEIDGLISYLSISLNKNSIYVEELSVDNPNFTISEVIR
NYSYLDNSYROSSLSKLYKNYOLYNYFFETSLYEVDNKNKSYSLSKNTDGINI
SSVFKLINDKESYVQRWDECIICVLDGTEYLDISPENNRICIVLSKDNKAKITV
NTDLFRPCDITFSYNDKRYFESLSLDGSDYNNMICNDNNKVPKGAHLWILES"
7931..12032
/gene="BONT/C1"
/mrna ">12032
7931..>12032
/gene="BONT/C1"
RBS 8016..8021
/gene="BONT/C1"
CDS 8032..11907
/gene="BONT/C1"
/codon_start=1
/transl_table=11
/protein_id="CA51313.1"
/db_xref="GI:516175"
/db_xref="swiss-prot:P18640"
/translation="MPTIINFNYSQVDNKNILYLDTHLNLANPEKAFRITGNIW
VIPFRSRNPNLNKIPRVTSPKSGYIDPNYLSDDKDTFLKEIKLFRKINSREI
GEELIYRLSDTPEPGNNTPINTFDFDVSVDVKTQGNWVKTGINSVLTIG
PRENIIDPESTFKLTNTFAQEGGALSISISPREMLTYSNADNDYEGEPKSE
FCMDPILILHNLHMHENLYGAIENDQRISSVTSNIPYSQYNKLEVAEYARGGP
TILIPKASRYFEKALDIYRSIAKLNSIITANPSPNKYISGVTQKQLIKRYFPV
ESSEGVIVNRKFYNEIQTFTENYAKIYNQNRKIYSNVTPTANILDNDY
YDQNGNFKPSNLNLFMGONLSRNPALRKNENMLYLFKFKHAIDGRSLYNT
LDRGELVNTDLPFIDISDKTDFIPLRKIDINEETEVIYPDNVSDOVILSKNTE
HGOLDLVPDSSEILPENOVFYDNRTONYDLYNSVYLESQKLSDNBDFPTR
SIEBALONSADVTFEPLANKYNAGVGGGLFELMWANDVVEDEFTNLRKDLKSID
VSAIPIGPAINISNVRSGNTEAPAVGVVILLEAPEFTIPALGAFVYSKVOE
RNEIILKIDNLEGRNKRKDSIENMGMTWLSRIITOPNNISQMDOSLNYQAGAIKA
KIDLEYKISGSDENKSOVENLKLSDVLIKSEAMNNKIFRESVSLYLFKNMLPK
VIDELNEFRNTKLNLDHNLILVGEVDKLKAKVNNSFQNTIPFIFSYTNNLS
LKDINFEYNNIDKLSLQNRKNTLVDTSYNAEVSDEGVDQLNPFPDFKLGSS
IDGKLVITQENIVNYSVMVESFISFWIRINKVSNLPGYTIDSVKNNSGWSIGI
ISNLFVTLKQENDESINFSYDINSNAPKNKFWFTVTNNMKNMKIYNGKLIID
TIKVKELGINFSTJFEINKLPDGLITSDSDNINMWLRDFYIFAKELDGKDLIL
FNSLQYNVYKDWGNLDRYNKYVMYNDIYLRNRYANSRQIVFNRRNNDNFEGY
KIIIRKRGNTNDRVRGGDILYPMNTINNKAYNLFMKNMETIADNHESTDIYGLR
EOTKDINDNIIOQPNNTYYASQIFKSNFNGENISIGSISGTYFRFLRGDWDYRHN
YLVTYKQNTVYASLESSTSHWGFVPVSE"
11993..12032
/terminator
CASE COUNT 4825 a 1287 c 1703 g 4482 t
ORIGIN
Query Watch 48.5%; Score 665.2; DB 7; Length 12297;
Best Local Similarity 68.3%; Pred. No. 1.4e-92;
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;
13 ACCATCCCATCAACATCTCTCCACACCACTCCCTGTTGAAGGACATCATCAAC 72
10558 ACAATACCTTTAATATTTTTCATATCTAATAATATCTTTTATTAAGATATAAT 10617
73 GAGTACTTCAACAACATCAACGACTCCAGATCCGTCCCTGCGAGAACCCGPAAGAAC 132
10618 GAATATTTCAATAATATTATGATCAAAAATTTTGAGCCTACAAAACAGAAAAAATACT 10677
133 TTGGTCGACACTCCGGTTACACGCCGAGGTCCTCCGAGGAGGTGAGCTCCAGCTGAAC 192
10678 TTAGTGGATACATCAGGATATTAATCGAAGTGAAGTGAAGGAGGATGTTTCAGCTTAAT 10737
193 CCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCCTCCGGTGAGGACAGAGGTGAAGTCATC 252
10738 CCAATATTTCCATTTGACTTTAAATAGTAGTCTCAGCGGAGGATAGAGGTAAAGTTATA 10797
253 GTACCCAGACAGACATCGCTACACTCCATGCTAGGAGTCTCTTCCATCTCCTTC 312
10798 GTAACCCAGAGTAAATAATTTGTATATATCTATGTAAGAAATTTTAGCATTAGTTT 10857

```

```

QY 313 TGGATCAGAATCAACAAGTGGGTCTCCAACTTGCAGGTTACACCATCATCGACTCGGTC 372
Db 10858 TGGATCAGAATCAACAAGTGGTGAAGTAAATTTACCTGGAATATACTATAATGATGTT 10917
QY 373 AGAACAACACTCGGTTGGTCCATCGGTATCATCTCCAACCTCCCTGGTCTTCAACCTGAAG 432
Db 10918 AAAATAAATCAAGTGGTGAAGTATAGGTATATTAGTAAATTTTTTAGTATTACTTTAAAA 10977
QY 433 CAGAACGAGGACTCCGAGCAGCTCCATCAACTTCTCCTACGACATCTCCAAACAACGCTCT 492
Db 10978 CAAATGAAGTAGTGACAAAGTATAAATTTTAGTTATGATATCAATAATGCTCT 11037
QY 493 GGTACACAACAGTGGTCTTCGTCCACCGTCAACAACATGATGGTGAATCAAGATC 552
Db 11038 GGATACAATAAATGGTTTTTGTAACTGTTACTAACAATATGATGGGAAATATGAAGATT 11097
QY 553 TACATCAACGGTGAAGTGCATCGACACCATCAAGTCAAGGTCAAGGATTCAGCGTATCAACTTC 612
Db 11098 TATATAATGGAATTAATTAAGTACTATTAAGTAAAGTAAAGTAACTGGAATTAATTTT 11157
QY 613 TCCAAAGCATTCACTCGAGATCAACAAGATCCAGACACCGGTCTGATCACTCCGAC 672
Db 11158 AGCAAACTATAAATTTGAATAAATAAATAAATTCAGATACCGGTTTGTACTTACAGAT 11217
QY 673 TCCGACACATCAACATCTGATCGGTGACTTCTACATCTTCGCCAAGGAGTTGGACGGT 732
Db 11218 TCTGATACATCAATGTTGGTGAAGATTTTATATTTTGTCTAAGATTTAGATGGT 11277
QY 733 AAGSACATCAACATCTCTCAACTCTTCGAGTACACCAAGTCTGCAAGGACTACTG 792
Db 11278 AAGATATTAATATATTTTAAATAGCTTGCATATATACTAATGTTGTAAAGATTTATGG 11337
QY 793 GGTACGACCTGAGATCAACAAGGACTACTACATGCTCAACATCGACTACTTGAACAGA 852
Db 11338 GGAATGATTTAAGATATAAAGATATTTATGTTTATATAGTATATAGATTTATTAATAGA 11397
QY 853 TACATGTACGCAACTCCAGACAGATCTCTTCAACACAGACAGCTTCAACAACAGACTTC 912
Db 11398 TATATGTGCAACTCAGACAAATTTCTTTTAACACAGTACAGAAATAATAATGACTTC 11457
QY 913 AAGGAGGTTACAAGATCATCATCAAGGTATCAAGGTATCAGAGGTACACCAACGACAGAGTC 972
Db 11458 NATGAAGGATATAAATAATTAATAAAGATCAGAGAAATACAAATGATAGTAGATA 11517
QY 973 AGAGTGTGTACATCTCTGACTTCTGACATGACTTCAACAACAAGGCTTCAACCTGCTTC 1032
Db 11518 CGAGGAGGAGATATTTTATATTTTATGATGACAATTAATAACAAGCATATAATTTGTTT 11577
QY 1033 ATGAACAGGACCATGTAGCCGACAAACACTCCACCGAGGACATCTAGCCCATCGGT 1092
Db 11578 ATGAAGATGAACATGTATGTCAGATATATCATAGTACTGAAGATATATATGCTATAGGT 11637
QY 1093 CTGCGTGAACAGCAAGGACATCAACGACAAACATCATCTTCCAGATCCAGCCCAATGAAC 1152
Db 11638 TTAAGAGACAAACAAAGGATATAATGATAATATATATATTTTCAATACAACCAATGAAT 11697
QY 1153 ACACHTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAGAGACATCTCC 1212
Db 11698 AATACCTATTATTACGCATCTCAAAATATTAAATCAAAATTTTAATGAGGAAATATTTCT 11757
QY 1213 GGTATCTGTTCATCGGTACCTTACAGATTTCCGTCTGGGTGCTGACTGCTACAGACACAAC 1272
Db 11758 GGAATATGTTCAATAGGTACTTATCGTTTAGACTTGGAGGTGATGTTATAGACACAT 11817
QY 1273 TACTTGGTTCACACTGTCAAGGAGTAACTACGCTCTCTTCTGCTGGAGTCCACTTCCACC 1332
Db 11818 TATTTGCTGCCCTACTGTGAAGCAAGGAAATATGCTTCATTATTAGAATCAACATCACT 11877
QY 1333 CACTGGGAGTTCGCTCCAGTCTCCGAGTAA 1362
Db 11878 CATGGGGTTTTGTACCTGTAAGTGAATAA 11907

```



```

RESULT 8
AB037920      11631 bp      DNA      linear      BCT 24-JAN-2002
LOCUS         Clostridium botulinum genes for ORF-22, HA-70, HA-17, HA-33, NTNHA,
DEFINITION    neurotoxin, complete cds.
ACCESSION     AB037920
VERSION       GI:6939789
KEYWORDS      neurotoxin; NTNHA; HA-33; HA-17; HA-70; ORF-22.
SOURCE        Clostridium botulinum (strain:D-4947) DNA.
ORGANISM      Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
               Clostridium.
REFERENCE     1
AUTHORS       Kouguchi,H., Watanabe,T., Sagane,Y., Sunagawa,H. and Ohyama,T.
TITLE         In vitro reconstitution of the Clostridium botulinum type D
               progenitor toxin
JOURNAL       J. Biol. Chem. 277 (4), 2650-2656 (2002);
MEDLINE       21659747
REFERENCE     2 (bases 1 to 11631)
AUTHORS       Sagane,Y., Watanabe,T., Kouguchi,H., Yamamoto,T., Takizawa,J.,
               Kawabe,T., Murakami,F., Muroga,A., Nakatsuka,M. and Ohyama,T.
TITLE         Direct Submission
JOURNAL       Submitted (01-FEB-2000) Tohru Ohyama, Tokyo University of
               Agriculture, Faculty of Bioindustry; Kasaka 196, Abashiri, Hokkaido
               099-2493, Japan (E-mail:t-oyama@bioindustry.nodai.ac.jp,
               Tel.81-152-48-3638(ex.368), Fax:81-152-48-2940)
FEATURES             Location/Qualifiers
     source          1..11631
                     /organism="Clostridium botulinum"
                     /strain="D-4947"
                     /db_xref="taxon:1491"
     gene            6..545
                     /gene="orf-22"
     CDS             6..545
                     /gene="orf-22"
                     /codon_start=1
                     /transl_table=11
                     /product="ORF-22"
                     /protein_id="BAA90656.1"
                     /db_xref="GI:6939790"
                     /translation="MNDLFAENLKHNDHDFDIEMLSKKYIEKTKYNYLYOYNN
                     DILHKKELNEINSELDLRYKISIRYCINICKNRDKRIIYSEATYK
                     IRANVYSCEDEFELDTLSLNYKEQIIVKFFECRKNQETARRLHLRSQSYK
                     IRIKSLKLYIVNQLVNI"
     complement(621..2492)
     /gene="ha-70"
     complement(621..2492)
     /gene="ha-70"
     /codon_start=1
     /transl_table=11
     /product="HA-70"
     /db_xref="GI:6939791"
     /translation="MSLSIKELVYTKDSTNNVNLADGVVYVNRGSGWILSRQNMNIG
     GNISNCTATVGLRIREFATPYVPTAFSENEYIRNNVQVFAFTEASRPIGFE
     FSKTAPSNKGYNLQTYIRYELIKVLRTVIERAVLVPSLGYAKSFEFSGDID
     ANFTSDXCLNEKFTYKRIKIAFTTAKESDSDNNTNTNTSQTILPYPNGLYVINK
     GDGYNRNDKILLETITSGSIQPRLTRPLTNTSPTLFSQYEYEARLND
     AFNQLQNTSTLTFVEEAPDNKNSKMAKNTYKELINTQNGNADKAEYILPSL
     GKCEVSAPQAPVPTVQDGFQIOTGNENIIVGVINPSENIEISTPIPDYTY
     NPTSDONNACVLTFTVNTGVKINAONNLPPLIIVESIGSDNMNQISNTNNIK
     ALNITGDSSASVSLIVSLKKNKYIRIPIQSSSTNOLIFKELGNISDLANST
     VNILDNLNTSGTHYTCQSDPVGVNYSQITIPGDNNNIASSIFSTFRNNQIGTLY
     RLTESNGYNTITKNTSDLLNVPESLNGATYIRVYKVTNLNINIIFDAIRNS"
     complement(2506..2946)
     /gene="ha-17"
     complement(2506..2946)
     /gene="ha-17"
     /codon_start=1
     /transl_table=11
     /product="HA-17"
     /protein_id="BAA90658.1"
     /db_xref="GI:6939792"

/translation="MSSERTFLPNGYKIKSLFSDSLYLYTSSGSLFNTSSLDNQK
WKLEYSSNGFRFNSVAPNPKYLAINDYGFYVLSSSNNSLWPKIKAINSYIICTL
SIVNVDYANTIVDNNNTDQILLNPNFDINNNSQILKLEK"
complement(3008..3868)
/gene="ha-33"
complement(3008..3868)
/gene="ha-33"
/codon_start=1
/transl_table=11
/product="HA-33"
/protein_id="BAA90659.1"
/db_xref="GI:6939793"
/translation="MSQTNANDLRNNEVEFFISPSNNTNKVLKDISOSEVKLNKLSGA
NQWRLYIDNKQAYTIKVMNDTSLTLTWDAPLSSVSVAITDNTNNNQIWLQDTRISR
NVLIRNMPNLVQLNTDITLIVTQTSNNQFFKFSNCIYEALNRNCKIQTOLNS
DRFLSKNLNSQIIVLQWDFDSSRQWTIEYNETKSAYTLKQENRNYLFWGNSNYV
ETVQSDTSLQYWNINYLONDASKYILYNLQDTRVLDVYNSQTANGTHVIVDSYHGN
TNOQWIINLI"
4129..7719
/gene="ntnha"
4129..7719
/gene="ntnha"
/codon_start=1
/transl_table=11
/product="NTNHA"
/protein_id="BAA90660.1"
/db_xref="GI:6939794"
/translation="MDINDLNINSVDNKNVIVFAKNTTFFKAKVAPNIWAVE
RYGEPDLIAEYKLDGGYDSNLSQDSERENFLOAIITLKRNNNTISGKLSLLI
STAIPFYGVGGYSPNITFTGKTPKSNKLNLSVTITPFFGQVRETNRYESQN
KNFYASNIIVIFGPNIVENNVICVKKNDAGMGTMALIFQLLYKYKFTYIDP
AMELTKLISLFLYKINPDDLVVPLRTLONKQFSLNIDLSISGVDLEFI
NTNPYFTNSYFSNITKPEKYNIVETIEGNAIGNDIKLRLKOKTQNSVODLWNL
NLNFSKEPNSIIPDRFSNALKHFYKQYTYDYGDNVINGVANGOINTKLPLSDKN
TNIISKEPNSVNLVANNISLAKSNISYIGDLKGTTEDFYSTKILPNEYEYRNDSD
NFPLNNTISEVDSIPEIDINPYKDSDDLFTQITSTEEVITHTALPYNVLQAI
ITNENFTLSDFSKVWSKDSLVYSLDNLMSYLETIKNDGPIOTDKKYLWLKEVF
KNYSFINDLTQEDISGQINEVWIFGKALNLTNSFEYQNSGPISLSKKDLNF
SEPNEIDDIPLDLSLGLSKFNLNKLVEYISKNRVYFRVYFNFLDQWTEYSQYFE
LIMAKOSILAOESVVKQIIONKFTLSKASIPDPTKLKIKETTEKTFIDLSNESOIS
NRYVDNLEKASICVEVDIYKFIYSMEKYINNIKTREIORTCNTINDNEKSIIL
NSYTKIDLPKLIQAIKKNFNSQVQVQWEMLSYQALLPATRGPNSNILEDISGK
NTLIQYTESVELVGVNGESLYLKSNEVEISFNSNFTINGLNNTFICFLWFTGDD
DKTLKGNVNCNCGWEIYFDNGLVEIISDNGQESVYLSLVNINNNTYIISVDRL
KQDILLINDKNVANVSIQIILNYSIVNLSVKNNSYVEELSVDLPKPAVEVIR
NYFSYLDNSYIRDSKSLLEYKNYQLYNVFPETSLYEVDNNSKLSLKNRTDGINI
PSVKEKLINDESKGVQKWDECIICVSDGTEKYLDISPENNRIOLVSSKNNAKITV
NTDLPRPDCITFSYNDKYFSLSLRDGDYNMCMICNDNKNYKGAHLWLKAS"
7734..11591
/gene="nt"
7734..11591
/gene="nt"
/codon_start=1
/product="neurotoxin"
/protein_id="BAA90661.1"
/db_xref="GI:6939795"
/translation="MTWPKDFNSYDPVNDNDILYRIPQNKLIITPVKAFMITONINW
VIPERFSDSNLSKPPRPTSKYQYDPSYLSLTDQKDTFLKGLIKLFRKINERDI
GKGLINLYVGVSPFMDSDSTPDEFTDTRHTTNIAVEKFGESKWTNITPESLIFG
PCMLIDYLTASLTQGOCSNPSPGEGFTLSLKVAPEFLTFDSVTNOSSAVLAKSI
PCMDPVATLAMELHETSLHOLYGINIPSDAKRIPQVSEGFSDQGNVQPEELYTFGGS
DVEILPQIELQRLKALGHYKIAKLANINKTIPSSWSNIDKYKIFSEKYNFDK
NTIINGFNLTIKGFNIENSQNIERNPALQKLSSESVDLTFYKVLRLNRSRDSSTC
IQVKNNTLPYADKDSISQIFESQITDTEVNTSDFNSLDESIDLRKAVITNPEAR
DPLLPNTVMEFLNVPGEVEFYDDITKDVDYLSYIYLAQKLSNNVENITLTVVEE
ALGYSNKIYTFPLSLAKVNGKQAGFLNNAVEVEFTNIMKKDITDKTSDVSVAI
IPYIGPALNTGNSALRGNKQAFATAGVAFLEGGPEFTIPALGVFTFSIQEREKI
IKTIENCLEQVRKWDYSQMWNSLRTTTFNHSIYQWVDSLSYQADAJAKIDL
EYKYSQSDRENKSKQVENKSLDKISEAMNNINKFREGSVTYLFWMLPKVIDE
LNKREDLKTETELNIDSHNIIIVGEVDRLLKAKYNSFENFTIPFNISFTNNSLKDM
INEYFNSINDSKILSLQNKNTLMDTSGYNAEVRVEGNVOLNPIPFDFPKLGSDDDR

```



GKIVITQNERIVNAMYSEFSISFIRINKWYNLPYTIIDSVKNNSSWSIGIISNF  
 LVLTKQNSQDIINFSYDINKAAGYKWFVITITNMGNMIIYINGKILIDIKV  
 KELTGINFSTITFOMNKIDYSDSDNTNMWIDFYIFAKELDDKIDNILENSL  
 QYTNVVDYNGNLDYKVEYMINVYKNGKGVIVTRKNNDFNKGKII  
 KRIGTNDYRGENVLFVNTIDNKQYSLGWYKPSRLGTDLYPLGLDQDEIR  
 KYGFIIPONTEDYTASQFLPSNATIRLGLISGSYFKLGGDYNFNHBYLIPVI  
 KIEHYALLESTTHWVFPASB\*  
 BASE COUNT 4493 a 1254 c 1657 g 4227 t  
 ORIGIN

Query Match 32.8%; Score 449.8; DB 1; Length 11631;  
 Best Local Similarity 59.38; Pred No. 1.3e-59;  
 Matches 807; Conservative 0; Mismatches 542; Indels 12; Gaps 2;

QY	13	ACCATCCATCAACATCTTCTCTACACCAACACCTCCCTGTTGAAGGACATCAAC	72
Db	10248	ACAATACCCCTTAATATTTTTCATATACTAATAATCTTTATTAAGAATATGATTAAT	10307
QY	73	GAGTACTTCAACATCAACAGCTCCAGATCTGTCCTGCGAGACCGCTAAGACACC	132
Db	10308	GAATATTTCAATAGTATTAATGATTCAAAATTTTGAGCTTACAAAATAAAAAATACT	10367
QY	133	TTGGTGCACCTCCGGTTACACCGGAGCTCCCGAGAGGCTGACGTCACGCTGAAC	192
Db	10368	TTGATGGATACATCAGGATATAACGCGAAGTGAGAGTAGAAGGCAATGTTACGCTTAAT	10427
QY	193	CCAACTCTCCCATTCGACTTCAAGCTGGTTCCTCCGGTGAGGACAGCTAAGGTCATC	252
Db	10428	CCAAATTTCCATTTGACTTTAAATAGGTAGTTCAGGGGATGATAGAGGTAAGTTATA	10487
QY	253	GTCACCCAGAGAGACATCGCTTACAACATCCATGATGACGAGTCTTCTCCATCTCCTC	312
Db	10488	GTAACCCAGATGAATATTTGATATATGCTATGATGAAGTTTTAGTATTAGTTTT	10547
QY	313	TGGATCAGATCAACAGTGGTCTCCAACTTCCAGGTTCACACATCATCGACTCCGTC	372
Db	10548	TGGATTAGGATAAATAATGGTGAAGTAATTTACCTGGATATACATAATGTAGTGT	10607
QY	373	AAGAACAACCTCCGGTTGCTCGATCGGTATCATCTCCAACTCCCTGGTTCACCCGTAAG	432
Db	10608	AAAAATACTCAGGTGGAGTATAGTATTTAGTAAATTTTGGTGTACTTTAAAA	10667
QY	433	CAGACAGGAGCTCCGAGGATCCATCACTTCCTTACGACATCTCCACACAGCTCCCT	492
Db	10668	CAAAATGAATPAGTGACACAGATATAAATTTAGTATGATATATCAAAAGATGCTGG	10727
QY	493	GGTTACAACAAGTGGTCTCGTCCAGCTCCACCAACATGATGGTGAACATGAAGATC	552
Db	10728	GGATATATATATGCTTTTGTATCTATTACTACCAATATGATGGGAATATGATGAT	10787
QY	553	TACATCAACGGTAAGCTGATGACACCATCAAGGTCAAGGATGACCGGTATCACTTC	612
Db	10788	TATATAATGGAAATTAATAGTACTATAAAGTTAAAGAGTTAACTGGAATTAATTT	10847
QY	613	TCCAAGACCATCACCTTCGAGATCAACAGATCCAGACACCGGCTGATCACCTCCGAC	672
Db	10848	AGCAAACTATACATTTTCATGAATGAATAATTCAAATCTGCTTAATACCTCAGAT	10907
QY	673	TCCGACACATCAACATGATGGATCGGTGACTTCTACATCTTCGCCCAAGAGGTGGAGGT	732
Db	10908	TCTGATACATCAATATGATGGATAGGGATTTTATATCTTTGCTAAGAAATTAGATGAT	10967
QY	733	AGGACATCAACATCTGCTCAACTCTCTGAGTACACACAGCTCGTCAAGGACTACTGG	792
Db	10968	AAAGATATTAATATATTTTAAATAGCTTGCAATATCTACTATGTTGTAAGAAATATGG	11027
QY	793	GTTAAGGACTGAGATACACAAAGGAGTACTACATGTTGGTCAACATCTACTTGAACAGA	852
Db	11028	GGAAATGATTTAAGATATGATAAAGATATTACATGATTACGTAATAATATGATGATA	11087
QY	853	TACATCTAGCCCACTCCAGACAGATCTCTTCAACACACAGCTTACACACAGCTTC	912
Db	11088	TATATGCTCTAAAAAAGGCAATGGAAATTTGTTTTTATACACGCTAAAAATAATGACTTC	11147

Copied from 10910186 on 05-05-2004

QY	913	AACGAGGTTTAAAGATCATCATCAACGCTATCAGAGGTAAACCAACGACACGAGCT	972
Db	11148	AATGAAGGATATAAATATTAATAAAGAAATAGAGGAAATCAAAATGATAGTACAGTA	11207
QY	973	AGAGTGGTGACATCTCTGCTACTTCGACATGACATCAACACAGGCTTACACCTGTC	1032
Db	11208	CGAGAGAAATGATATATATTTTAATACATCAATGATACAAACAATATAGTTAGGT	11267
QY	1033	ATGAGAACGAGACATGATAGCGGACAAACCACTCCACCGAGGACATCTACGCCATCGT	1092
Db	11268	ATGATATAACCTTCTA-----GAAATCTAGGAGTCTATTAGTTCCACTAGGTGCA	11318
QY	1093	CTGGGTGAGCAGACAGGACATCAACGACAAACATCATCTCCAGATCCACGCAACGAAC	1152
Db	11319	TTGGATCAACCAATGGATGAGATAGCTAAATATGGTTCTGTTTATAATACAACTGCAAT	11378
QY	1153	AACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGACATCTCC	1212
Db	11379	ACTTTGATTACTATGCATCACTATTTTGTCAAGTATGCAACAAACAAATAGGCTT	11438
QY	1213	GTATCTCTTCCATCGGACCTACAGATTCGGTCTGGTGGTGAC---TGTACACACAC	1269
Db	11439	GGAATPACTATCAATTTGGTATAGTTTCAACCTTGGAGATGACTATTTGGTTAATC	11498
QY	1270	AACTACTTGGTTCCTCAACTGTCAACGAGGTAACTACGCTCTTGGTGGAGTCCACTTC	1329
Db	11499	GAATATTTAATCTCTGTATATAAATAGAGCATTTATCTTCATTTATAGATCAACATCA	11558
QY	1330	ACCCACTGGGATTCGTCCTCCAGTCTCCGAGTATAGGAAT 1370	
Db	11559	ACTCATTTGGTTTTTGTACCTGCAAGTGAATAATAAATAGT 11599	
LOCUS	CLOTNCO 4049 bp DNA linear B01 18-DEC-2001		
DEFINITION	Clostridium botulinum type D gene for neurotoxin consisting of botulinum neurotoxin D and C1, complete cds.		
ACCESSION	D38442		
VERSION	D38442.1 GI:1374775		
KEYWORDS	neurotoxin consisting of botulinum neurotoxin D and C1; neurotoxin.		
SOURCE	Clostridium botulinum D (strain:South African) DNA.		
ORGANISM	Clostridium botulinum D		
REFERENCE	1 Moriishi, K., Koura, M., Abe, N., Fujii, N., Fujinaga, Y., Inoue, K. and Ogumad, K. Mosaic structures of neurotoxins produced from Clostridium botulinum types C and D organisms. Biochim. Biophys. Acta 1307 (2), 123-126 (1996)		
AUTHORS	Moriishi, K., Koura, M., Abe, N., Fujii, N., Fujinaga, Y., Inoue, K. and Ogumad, K.		
TITLE	Mosaic structures of neurotoxins produced from Clostridium botulinum types C and D organisms		
JOURNAL	Biochim. Biophys. Acta 1307 (2), 123-126 (1996)		
MEDLINE	96283801		
REFERENCE	2 (bases 1 to 4049)		
AUTHORS	Moriishi, K.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-SEP-1994) Kohji Moriishi, Osaka University, Research Institute for Microbial Diseases, Research Center for Emerging Infectious Diseases; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:kohji@biken.osaka-u.ac.jp, Tel:81-6-6879-8343, Fax:81-6-6879-8269)		
FEATURES	Location/Qualifiers		
source	1. 4049 /organism="Clostridium botulinum D" /strain="South African" /db_xref="taxon:36829" 163..4020 /transl_table=11 /product="neurotoxin consisting of botulinum neurotoxin D and C1" /protein_id="BA07477.1" /db_xref="GI:1374776"		
CDS			

/translation="MTWPKDFNYSDPVNDNDILYLRIPOKLIITPPVKAFMITONIW  
VIPERSSDNPSPRPRTSKYSDYPSYLSDEQDTEKLGILKFLKREINERDI  
GKLLIVVSGFMDSDPDETFDTRHTNIAVEKFEKSGWKNITITPSVLIFG  
PLNILDYASLQGGQSPFEGFTLSLKVADEFLLTDSFTSNQSSAVLGKSI  
FCMDPVIALHRELTSLHOLYGINPSDKRIPOVSEGFSDGPNVPOFERYLTGGS  
DVEIPIERLQRELAHGHYKDIARLNINKTIPSSWSNIDKVKIFSPKYNFKD  
DNUGFNLYNDKNSLYSULTNVMSVLYSSQYNKRNHYKSKHLPVFNILDDNI  
YIINGFNLTIGFNGNSQNLERNPALQKLSSESVDLFTKVLRLTRNRSDDSTC  
IOVKNLTPVADKDSIOEIPESQITDETENVYSDNLSDESILDAKVPNEPV  
DPLPNVMEPLANVPEEVEYDDITKDVLNSYLYLEAKLSNNVENITLTTVEE  
ALGYSNRYTFLPSLAENKVGQAGLELNWANEVDEFTNIMKDDILDKLSVSAI  
IPYIGPALNIGKAGNPKOATAGVAFLLGPEFTIPALGVTFVPSIQEBREKI  
IKTIGLEORVRWSDYOMYSNLSRITTOFNIHSYOMVDSLSYADAJAKIDL  
EYKYSDDENIKSQRNLSKNSLDVKISEAMNNINKFIRECSVYLEKNMPLPKVIDE  
LNKFDKIKELIINLQNKNTLMDVEYDLKAKYNSFENITPFNIFSTNNLSLKKM  
INFEFNSINDKLSLQNKNTLMDVEYDLKAKYNSFENITPFNIFSTNNLSLKKM  
GKIIVTQENIVNMYTESISFWIRKKNVNSLPGYITIDSVKNSGWSIGIISNF  
LVFTLKNSSEDDINFDYISKNAAGYKWFVTITNMGMNMIYINGKILDIKV  
KELTGINSFTIIPQNKIPNTGLITSDSNINWIRDFYIFAKELDDKIDNINLSL  
YOTVNDVYGNLRYDKYVIMNVNMYKSKGIVFTRKNNNDNBYKRII  
KIRGNITDVRVSENVLYPNTIDKQYSLGMYKPSNLGTDLPVGLADQPMDEIR  
KYGSFIQPCNTEDYASQFLSSNATNRLGILSIGSYSPKLGDDYFNHEYLPIVI  
KIEHFALESSTSHWVFVASE"

Query Match 32.7%; Score 448.2; DB 1; Length 4049;  
Best Local Similarity 59.2%; Pred. No. 2.6e-59;  
Matches 806; Conservative 0; Mismatches 543; Indels 12; Gaps 2;

13 ACATCCCATTCATCTTCTTACACCAACATCCCTGTTGAAGGACATCATCAAC 72  
2677 ACAATACCCCTTTAATATTTTCATATACTAATAATCTTTAATAAAGATATGATTAAT 2736  
73 GAGTACTTCAACACATCAAGACTCCCAAGATCCTGCTCCGTCAGAACCGTAAGAACACC 132  
2737 GAATATTTCAATAGTATTAAATGATTCAAAATTTTGACITACAAAATAAAAAATACT 2796  
133 TTGGTCGACCTCCGGTTACAAACCGGAGGTCTCCGAGGAGGTGACGCTCCAGCTGAAC 192  
2797 TTGATGATACATCAGATATAGATTAAGCAGAACTAGAGTAGAAGCAATGTTTCAGCTTAAT 2856  
193 CCAATCTCCCATTCGACTCAAGCTGGTCTCCGTCGAGGACAGGTAAAGGTCATC 252  
2857 CCAATATTCATTTGACTTTAAATAGTAGTAGTTCAGGGATGATAGAGTAAATATA 2916  
253 GTACCCAGACGAGAACATCGTCTACAACTCCATGACGAGTCTCTCCATCTCCTTC 312  
2917 GTAACCCAGATGAAATATTGTATATGCTATGATGAAAGCTTTTAGTATTAGTTT 2976  
313 TGGATCAGATCAACAGTGGTCTCCAACTTCCAGGTTACACCATCATCGACTCCGTC 372  
2977 TGGATAGGATGAAATAGTGGTAGTAAATTTACCTGGATATACATATAATGATAGTTT 3036  
373 AGAACAACCTCCGGTGGTCCATCGATATCATCTCCAACTTCCGTCCTTCCAGCTGAAG 432  
3037 AAAAAATCACTCAGTTGGAGTATAGTATTATTAGTAAATTTTGGTGTACTTTAAAA 3096  
433 CAGACAGGAGCTCCGAGGATCCATCACTTCTCCAGACATCTCCACACAGCTCCCT 492  
3097 CAAATGAAATAGTGACACAGATATAAATCTTATGATGATATATCAAGAATGCTGG 3156  
493 GGTACACAGAGTGGTCTTCCCTCCGCTCCACCAACATGATGGGTAAACATGAAGATC 552  
3157 GGATATAATAATGTTTCTTCTAATCTTACTACCAATATGATGGGAATATGATGATT 3216  
553 TACATCAAGGTAAGTATGATGACACATCAAGGTCAAGGAGGTGACCGGTATCAATTC 612  
3217 TATATAATGGAAATTAATAGATCTATAAAGTTTAAAGATTTAACTGGAAATTAATTTT 3276  
613 TCCACACCATCATCCTCGATGATCAACAGATCCAGACACCGGTCTGATCACCTCCGAC 672  
3277 AGCAAACTATACATTTCAATGATGATTAATTTCCAAATCTGCTTAATTAACCTCGAT 3336

QY 673 TCCGACACATCAACATCGTGGATCCGTCGACTTCTATCATCTTCCGCAAGGAGTGGACGGT 732  
Db 3337 TCTGATACATCAATATGATGATGAAGGATTTTATATCTTTGCTAAAGAAATAGATGAT 3396  
QY 733 AAGGACATCAATCCTGTTCAACTCTCTGACGTACACCAACGTCGTCAGGAGTACTTGG 792  
Db 3397 AAGATATTAATATATTATTAAATAGCTTGCATATATCTAATTTGTTAAAGATTTATGG 3456  
QY 793 GGTAAACGACCTGAGATCAACAAAGGATCTACTATCGTCAACATCGACTACTTGAACAG 852  
Db 3457 GGAATATGATTTAAGATATGATAAAGAAATATTACATGATTAACGTAATATATCAATAGA 3516  
QY 853 TACATGTACGCAACTCCAGACATCGTCTTCAACACGAGAGCTTACACACACGACTTTC 912  
Db 3517 TATATGCTTAAAGCAAGCAATGGAATTTTAAATACACGTAATAATAATATGACTTC 3576  
QY 913 AACGAGGTTTAAAGATCATCATCAACGCTATCAGAGGTAAACCAACGACACGAGTCC 972  
Db 3577 AATGAAGGATATAAATTTAATAAAGAAATTAGAGAAATACAAATGATAGTAGATA 3636  
QY 973 AGAGGTGTGACATCCTGTTCTGACATGACATCAACCAAGGCTTACACACCTGTTTC 1032  
Db 3637 CGAGGAGAAATGATTTAATTTAATCTACATTTGATACCAACAAATATATAGTTAGGT 3696  
QY 1033 ATGAGAACGAGACCATGTACGCGCAACACCTCCACGAGGACATCTACGCCATCGGT 1092  
Db 3697 ATGTATAAACCTTCTA-----GAAATCTAGGAGTGAATTTAGTTCACATAGGTGCA 3747  
QY 1093 CTGGGTGAGCAGACCCAGGACATCAACGACCAACATCATCTTCCAGATCCAGCAATGAAC 1152  
Db 3748 TTGGATCAACCAATGGATGAGATACGTAATATGTTGTTTATATACCACTGATTTAT 3807  
QY 1153 AACACTTACTACTAGGTTCCAGATCTTCAAGTCCAACTTCAACGCTGAGAGACATCTCC 1212  
Db 3808 ACTTTTGATTACTATGATCATCACAATTTTGTGCAAGTAATGACAAACAAATAGGCTT 3867  
QY 1213 GGTATCTGTTCCATCGTACCTACAGATTCGGTCTGGTGGTGAC---TGTGACAGACAC 1269  
Db 3868 GGAATACATCAATGGTGGTATGATTTTCAAACTTGGAGATGACTATTGGTTTATACAC 3927  
QY 1270 AACTACTTGGTGGTCAACCTGTCACAGGAGTAACTACGCTTCTGTTGGAGTCCACTTCC 1329  
Db 3928 GAATATTTAATCTGTTATATAAATAAGAGCATTTATCTTCATTATTAGATCAACATCA 3987  
QY 1330 ACCCACTGGGATTCGTCGCCAGTCTCCGAGTATATAGGAAT 1370  
Db 3988 ACTCATTTGGTGGTGGTACCTGCAAGTGAATAATAATTAAGT 4028

RESULT 10  
AC117900/c  
LOCUS AC117900 151705 bp DNA linear HTG 18-JUL-2002  
DEFINITION Rattus norvegicus clone CH230-320C16, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 33 unordered pieces.  
ACCESSION AC117900  
VERSION AC117900.3 GI:21746813  
KEYWORDS HTG: HTGS PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 151705)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Avele,M., Banks,T.,  
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyte,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,

Copied from 10910186 on 05-05-2004



Query Match	3.7%	Score 198.4;	DB 2;	Length 176351;
Best Local Similarity	49.6%;	Firm. No. 9.3e-20;		
Matches 483;	Conservative 0;	Mismatches 491;	Indels 0;	Gaps 0;
236 ACAGAGTGAAGGTCATCGTCACCCAGAACGAGCAATCGTGTACAACTCCATGTACGAGT 295				
52708 ACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 52649				
296 CATTCTCCATCTCCTTCTTGATCGAATCAACAAGTGGGTCTCCAACTTGCCAGGTTACA 355				
52648 ACAACAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 52589				
356 CCATCATCGACTCGGTACAGCAACTCCGGTGTGCTCCATCGGTATCATCTCCAACTCC 415				
52588 GCAGCAACAACAACAACAGCAACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 52529				
416 TGGTCTTCCACCTTGAGAGAGAGAGGACTCGGAGCAGTCCATCAACTTCTCCACGACA 475				



\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1195: contig of 1195 bp in length
* 1196: gap of unknown length
* 1296: contig of 1006 bp in length
* 2302: gap of unknown length
* 2402: contig of 1196 bp in length
* 3598: gap of unknown length
* 3698: contig of 1008 bp in length
* 4706: gap of unknown length
* 4806: contig of 1248 bp in length
* 6054: gap of unknown length
* 6154: contig of 1556 bp in length
* 7110: gap of unknown length
* 7810: contig of 1286 bp in length
* 9076: gap of unknown length
* 9176: contig of 1212 bp in length
* 10387: contig of 1212 bp in length
* 10488: contig of 1684 bp in length
* 12172: gap of unknown length
* 12272: contig of 1852 bp in length
* 13524: gap of unknown length
* 14024: contig of 1230 bp in length
* 15254: gap of unknown length
* 15354: contig of 1703 bp in length
* 17057: gap of unknown length
* 17157: contig of 1474 bp in length
* 18631: gap of unknown length
* 18731: contig of 1834 bp in length
* 20565: gap of unknown length
* 20665: contig of 2256 bp in length
* 22921: gap of unknown length
* 23021: contig of 1022 bp in length
* 24042: gap of unknown length
* 24143: contig of 1265 bp in length
* 25408: gap of unknown length
* 25508: contig of 1819 bp in length
* 27327: gap of unknown length
* 27427: contig of 1399 bp in length
* 28826: gap of unknown length
* 28926: contig of 1370 bp in length
* 30296: gap of unknown length
* 30396: contig of 2506 bp in length
* 32902: gap of unknown length
* 33002: contig of 2382 bp in length
* 35384: gap of unknown length
* 35484: contig of 2793 bp in length
* 38276: gap of unknown length
* 38376: contig of 1492 bp in length
* 38777: gap of unknown length
* 39869: contig of 2001 bp in length
* 39969: gap of unknown length
* 41970: contig of 2313 bp in length
* 42070: gap of unknown length
* 44382: contig of 3110 bp in length
* 44383: gap of unknown length
* 44483: contig of 4459 bp in length
* 47593: gap of unknown length
* 47693: contig of 4509 bp in length
* 52152: gap of unknown length
* 52252: contig of 4509 bp in length
* 56761: gap of unknown length
* 56861: contig of 4317 bp in length
* 61177: gap of unknown length
* 61278: contig of 4871 bp in length
* 66149: gap of unknown length
* 66249: contig of 5638 bp in length
* 71887: gap of unknown length
* 71987: contig of 6093 bp in length
* 78080: gap of unknown length
* 81154: contig of 2975 bp in length
* 81254: gap of unknown length

```

```

* 81255 88990: contig of 7636 bp in length
* 88991: gap of unknown length
* 88991 97276: contig of 8286 bp in length
* 97277 97376: gap of unknown length
* 97377 107376: contig of 10000 bp in length
* 107377 107476: gap of unknown length
* 107477 116754: contig of 9278 bp in length
* 116755 116854: gap of unknown length
* 118555 127329: contig of 10475 bp in length
* 127330 127429: gap of unknown length
* 127430 136746: contig of 9317 bp in length
* 136747 136846: gap of unknown length
* 136847 153947: contig of 17101 bp in length
* 153948 154047: gap of unknown length
* 154048 169163: contig of 15116 bp in length.
FEATURES
  location/Qualifiers
    1..169163
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-261G15"
BASE COUNT 41981 a 37857 c 38661 g 43240 t 7424 others
ORIGIN
Query Match 13.3%; Score 193; DB 2; Length 169163;
Best Local Similarity 50.6%; Pred. No. 6.3e-19;
Matches 469; Conservative 0; Mismatches 455; Indels 3; Gaps 1;
QY 233 AGACAGAGGTAAAGTTCATCGTCACCCAGACGACGACATCGTCTACAACTCCCATGTACG 292
DB 35132 AAGCCAAACCAACACACACAGCAGCAGCAACAAACAGCAACAGCAACAGCAGCA 35073
QY 293 AGTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 352
DB 35072 ACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 35013
QY 353 ACACCATCATCGACTCCGTCACGACAACTCCGTTGGTTCCTTCCTTCCTTCCTTCCT 412
DB 35012 ACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 34953
QY 413 TCCTGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 472
DB 34952 ACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 34893
QY 473 ACATTCCTCCAAACAGCTCTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 532
DB 34892 ACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 34833
QY 533 TGATGGCTTAACATGAAGATCTACATCAACGGTAACTGATCGACACCATCAAGTCAAG 592
DB 34832 GCAGCAACAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 34773
QY 593 AGTTGACCGGTATCACTTCCTTCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 652
DB 34772 ACAGCAACAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 34713
QY 653 CCGGTGTGATCACTTCCTTCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 712
DB 34712 ACAGCAACAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 34653
QY 713 TCCTTCAGAGGTTCGCGGTAGGACATCACTTCCTTCCTTCCTTCCTTCCTTCCTTC 769
DB 34652 GCACCAACAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 34593
QY 770 CCAACCTCGTCAAGGACTACTTCGCGGTAACTGAGATCAACAGAGGAGTACTACATGG 829
DB 34592 ACACCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 34533
QY 830 TCACATCGACTACTTGAACAGATACATGTAGCGCAACTCCAGACAGATCTCTTCACAC 889
DB 34532 ACAACCAACAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 34473
QY 890 CCAGAGCTTAACAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 949

```



Direct Submission  
Unpublished  
2 (bases 1 to 163034)  
Worley, K.C.  
Direct Submission  
Submitted (15-NOV-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 163034)  
Worley, K.C.  
Direct Submission  
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 11, 2002 this sequence version replaced gi:17973419.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
-----  
Center project name: GHWE  
Center clone name: CH230-103G20  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 79006 bases at least Q40  
Consensus quality: 84121 bases at least Q30  
Consensus quality: 88465 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/cenbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/cenbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 82 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1382: contig of 1382 bp in length  
\* 1383  
\* 1382: gap of unknown length  
\* 1383  
\* 2616: contig of 1134 bp in length  
\* 2617  
\* 2716: gap of unknown length  
\* 2717  
\* 4049: contig of 1333 bp in length  
\* 4050  
\* 4149: gap of unknown length  
\* 4150  
\* 5644: contig of 1495 bp in length  
\* 5645  
\* 5744: gap of unknown length  
\* 5745  
\* 6913: contig of 1169 bp in length  
\* 6914  
\* 7013: gap of unknown length  
\* 7014  
\* 8388: contig of 1375 bp in length  
\* 8389  
\* 8489: gap of unknown length  
\* 10139  
\* 10238: contig of 1650 bp in length  
\* 10239  
\* 10739: gap of unknown length  
\* 11735  
\* 11834: contig of 1496 bp in length  
\* 11835  
\* 13287: contig of 1453 bp in length  
\* 13288  
\* 13287: gap of unknown length  
\* 13388  
\* 14709: contig of 1322 bp in length  
\* 14710  
\* 14809: gap of unknown length  
\* 14810  
\* 15861: contig of 1052 bp in length  
\* 15862  
\* 15961: gap of unknown length  
\* 15962  
\* 17519  
\* 17618: contig of 1557 bp in length  
\* 17619  
\* 19300: gap of unknown length  
\* 19301  
\* 19400: contig of 1682 bp in length  
\* 19401  
\* 20829: gap of unknown length  
\* 20830  
\* 21929: contig of 1100 bp in length  
\* 21930  
\* 22029: gap of unknown length  
\* 22030  
\* 23315: contig of 1386 bp in length  
\* 23316  
\* 23416: gap of unknown length



```

* 23516 25034: contig of 1519 bp in length
* 25035 25134: gap of unknown length
* 25135 26213: contig of 1079 bp in length
* 26214 26313: gap of unknown length
* 26314 27705: contig of 1392 bp in length
* 27706 27805: gap of unknown length
* 27806 29246: contig of 1441 bp in length
* 29247 29346: gap of unknown length
* 29347 30523: contig of 1177 bp in length
* 30524 30623: gap of unknown length
* 30624 32166: contig of 1543 bp in length
* 32167 32266: gap of unknown length
* 32267 33674: contig of 1408 bp in length
* 33675 33774: gap of unknown length
* 33775 35023: contig of 1249 bp in length
* 35024 35123: gap of unknown length
* 35124 37010: contig of 1887 bp in length
* 37011 37110: gap of unknown length
* 37111 38878: contig of 1768 bp in length
* 38879 40329: gap of unknown length
* 40330 40429: gap of unknown length
* 40430 41896: contig of 1467 bp in length
* 41897 41996: gap of unknown length
* 41997 43475: contig of 1479 bp in length
* 43476 43575: gap of unknown length
* 43576 44684: contig of 1109 bp in length
* 44685 44784: gap of unknown length
* 44785 46637: contig of 1853 bp in length
* 46638 46737: gap of unknown length
* 46738 47987: contig of 1250 bp in length
* 47988 48087: gap of unknown length
* 48088 49255: contig of 1068 bp in length
* 49256 49355: gap of unknown length
* 49356 50589: contig of 1334 bp in length
* 50590 50689: gap of unknown length
* 50690 51988: contig of 1299 bp in length
* 51989 52088: gap of unknown length
* 52089 53966: contig of 1878 bp in length
* 53967 54066: gap of unknown length
* 54067 55332: contig of 1266 bp in length
* 55333 56379: gap of unknown length
* 56380 57079: gap of unknown length
* 57080 58450: contig of 1371 bp in length
* 58451 58550: gap of unknown length
* 58551 61372: contig of 2722 bp in length
* 61373 62639: contig of 1267 bp in length
* 62640 62739: gap of unknown length
* 62740 64180: contig of 1441 bp in length
* 64181 65566: contig of 1286 bp in length
* 65567 65666: gap of unknown length
* 65667 66959: contig of 1292 bp in length
* 66960 68733: contig of 1675 bp in length
* 68734 68833: gap of unknown length
* 68834 70507: contig of 1674 bp in length
* 70508 70607: gap of unknown length
* 70608 71880: contig of 1273 bp in length
* 71881 71980: gap of unknown length
* 71981 73275: contig of 1295 bp in length
* 73276 73375: gap of unknown length
* 73376 74950: contig of 1575 bp in length
* 74951 75050: gap of unknown length
* 75051 76776: contig of 1726 bp in length
* 76777 76876: gap of unknown length
* 76877 77949: contig of 1073 bp in length
* 77950 78049: gap of unknown length
* 78050 79783: contig of 1734 bp in length
* 79784 79884: gap of unknown length
* 82083: contig of 2200 bp in length

```

```

* 82084 82183: gap of unknown length
* 82184 83593: contig of 1410 bp in length

Query Match      13.1%; Score 180.2; DB 2; Length 163034;
Best Local Similarity 51.3%; Pred. No. 1.7e-18;
Matches 444; Conservative 0; Mismatches 418; Indels 3; Gaps 1;

QY 354 CACATCATCGTCCGTCGTCAGCAAGCACTCCGGTTGGTCCCATCGGTATCATCTCAACTT 413
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155250 CAACACAGCAGCAAGCAACAAACAAACAGCAACAGCAACAGCAACAGCAACAGGAA 155191
QY 414 CTTGGTCTTCAACCTGAAGCAGAGGAGCTCCGAGCAGTCCATCATCACTTCTCTACGA 473
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155190 CAACAACAACAGCAGCAACAGCAAGCAAGCAACAAACAGCAACAGCAACAGCAAA 155131
QY 474 CATCTCCACACAGCTCCTGGTTACAAACAAGTGGTTCCTCGTCCGCTCACAACAACAT 533
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155130 CAGCAACAACAGCAGCAGCAACAAACAGCAACAGCAACAGCAACAGCAACAGCAAA 155071
QY 534 GATGGGTAACATGAAGATCTAGATCAACCGTTAAGTGTGATCGACACCATCAAGGTCAAGGA 593
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155070 CAGCAACAACAGCAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAAA 155011
QY 594 GTTGACCGGTATCAACTTCTCCAAGCAATCACTTCAGATCAACAGATCCAGACAC 653
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155010 CAGCAACAATAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAA 154951
QY 654 CGGTCTGATCACTCCGACTCCGAGCAACATCAATGATGGATCCGTACTTCTACATCTT 713
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154950 CAACAACAGCAACAGCAACAAACAGCAACAGCAACAGCAACAGCAACAGCAAA 154891
QY 714 CGCAACAGGAGTTGGACGGTAAGGACATCAACATCTCTGTTCAACTCTTGGAGTACACAA 773
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154890 CAGCAACAACAACAGCAACAAACAGCAACAGCAACAGCAACAGCAACAGCAACAA 154831
QY 774 CGTCGTCAGGAGTACTCGGGTAACGACTGAGATACAAAGAGGAGTACTACATCGTCAA 833
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154830 CAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAA 154771
QY 834 CATCGACTTCTGACAGATACATGATACGCCAACTCCAGACAGATCGTCTTCAACACAG 893
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154770 CAGCAACAACAACAGCAACAGCAACAGCAACAGCAAA---CAGCAACAGCAACAGCA 154714
QY 894 ACGTAAACAACAGGAGTTCAAGCAGGGTTACAAAGATCATCATCAAGCGTATCAGAGTAA 953
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154713 CAACACACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCA 154654
QY 954 CACCAACGACACAGCAGTCAAGAGTGGTGACATCTCTGATTCGACATGACTATCAACAA 1013
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154653 CAGCAACAACAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAAC 154594
QY 1014 CAAGGCTACACCTGTTTCATGAAGACGAGAGCCATGATGCGCGACCACTCCACCGA 1073
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154593 CAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAG 154534
QY 1074 GGACATCTACGCCATCGGTCCTCGGTGAGCAGACCAAGGACATCAACAGCAACATCTT 1133
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154533 CAACAACAGCAACAGCAACAGCAACAGCAACAAACAGCAACAGCAACAGCAACAGCA 154474
QY 1134 CGAGATCCAGCCCAATGAACAACACTTACTACTACGTTCCCGATCTTCAAGTCCAAC 1193
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154473 CAACGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCA 154414
QY 1194 CAACGGTGAGAACATCTCCGGTATC 1218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154413 CAACGCAACAGCAACAGCAACAGCAAC 154389

RESULT 14
AC096212 165337 bp DNA linear HTG 11-JUL-2002
LOCUS Rattus norvegicus clone CH230-23N22, *** SEQUENCING IN PROGRESS
DEFINITION *** 68 unordered pieces.

```

AC096212 4 GI-21723349  
 HNG; HTGS\_PHASEL.  
 Norway rat.  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 165337)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C.,  
 Ausbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayale, M., Banks, T.,  
 Barabara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,  
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buhay, C., Burch, P., Burrell, C., Burrell, K.L., Byrd, N.C.,  
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Kocha, S., Durbin, K.J.,  
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B.,  
 Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratochvic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,  
 Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, F.,  
 Massey, E., Mathew, E., McLeod, M.P., Meador, M., Mei, G., Metker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokoko, S., Ogum, M., Okwuona, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojucan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,  
 Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, O.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 Worley, K.C.  
 2 (bases 1 to 165337)  
 Direct Submission  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 Worley, K.C.  
 Direct Submission  
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 10, 2002 this sequence version replaced gi:17943903.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: 3ERC  
 Center clone name: CH230-23N22  
 ----- Summary Statistics

Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 96557 bases at least Q40  
 Consensus quality: 103304 bases at least Q30  
 Consensus quality: 107916 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 68 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1	1019:	contig of 1019 bp in length
1020	1119:	gap of unknown length
1120	2182:	contig of 1063 bp in length
2183	2282:	gap of unknown length
2283	3490:	contig of 1208 bp in length
3491	3590:	gap of unknown length
3591	4592:	contig of 1002 bp in length
4593	4592:	gap of unknown length
4593	5977:	contig of 1285 bp in length
5978	6077:	gap of unknown length
6078	7186:	contig of 1109 bp in length
7187	7286:	gap of unknown length
7287	8413:	contig of 1127 bp in length
8414	8513:	gap of unknown length
8514	9692:	contig of 1169 bp in length
9693	9782:	gap of unknown length
9783	10917:	contig of 1135 bp in length
10918	11017:	gap of unknown length
11018	12051:	contig of 1034 bp in length
12052	12151:	gap of unknown length
12152	13374:	contig of 1223 bp in length
13375	13474:	gap of unknown length
13475	14588:	contig of 1114 bp in length
14589	14688:	gap of unknown length
14689	16062:	contig of 1374 bp in length
16063	16162:	gap of unknown length
16163	17290:	contig of 1128 bp in length
17291	17390:	gap of unknown length
17391	18879:	contig of 1489 bp in length
18880	18979:	gap of unknown length
20557	20557:	contig of 1578 bp in length
20558	20657:	gap of unknown length
20658	21994:	contig of 1337 bp in length
21995	22094:	gap of unknown length
22095	23618:	contig of 1524 bp in length
23619	23718:	gap of unknown length
23719	24845:	contig of 1127 bp in length
24846	24945:	gap of unknown length
24946	26433:	contig of 1488 bp in length
26434	26533:	gap of unknown length
26534	28392:	contig of 1859 bp in length
28393	28492:	gap of unknown length
28493	30382:	contig of 1890 bp in length
30383	30482:	gap of unknown length
30483	32223:	contig of 1741 bp in length
32224	32323:	gap of unknown length
32324	33841:	contig of 1518 bp in length
33842	33941:	gap of unknown length
33942	35473:	contig of 1532 bp in length
35474	35573:	gap of unknown length
35574	36870:	contig of 1297 bp in length
36871	36970:	gap of unknown length
36971	38600:	contig of 1630 bp in length
38601	38700:	gap of unknown length
38701	40028:	contig of 1328 bp in length
40029	40128:	gap of unknown length

Copied from 10910186 on 05-05-2004

```

Db 188 AATCAACATCGGTCTAAAGTTAACTTCGATCGGACAGAGATCAGATCCAGCTG- 246
Qy 221 GTTCTCCGGTGAGGACAGAGGTAAGGTATCGTACCCAGAGACAGATCGTCTACA 280
Db 247 -----TTCAATCTGGAATCTTCCAAATCGAAGTTATCCTGAAGAATGCTATCGTATACA 301
Qy 281 ACTCCATGATGAGTCCCTTCCCATCTCCTCTGGATCAGATCAACAAGTGGTCTCCA 340
Db 302 ACTCTATGATGAGAACTTCTCCACCTCTCTGGATCGGTATCCCGAATACTTCAACT 361
Qy 341 ACTTGCC-----AGGTTACACCATCATCGACTCCGTCACAGAACTCCGGTTGGT 391
Db 362 CCATCTCTCTGAACAATGAATACACCATCAATCAATGATGGAATCTGTGTGGA 421
Qy 392 CCATCGGTATCATCTCCAACTTCTCTGCTTCCCTTACCCCTGAGAGACAGGAGTCCGAGC 451
Db 422 AAGTATCTCTGAACATCGGTGAATCATCTGAGTCTGAGGACACTCAGGAAATCAAAC 481
Qy 452 AGTCCATCAACTTCTCTTACGACATCTCCAAACAGGTCTCTGTTTACAACA---GTGGT 508
Db 482 AGCGTGTGTTATCAAAATACCTCTCAGATGATCAATCTCTGACTACATCAATCGGTGA 541
Qy 509 TCTTCGTACCGGTACCAACAACATGATGGTAAATGATGATGATGATGATGATGATGATG 568
Db 542 TCTTCGTACCGGTACCAACAACATGATGGTAAATGATGATGATGATGATGATGATGATG 601
Qy 569 TGATCGACACCATCAAGGTCAAGGAGTGTGACCGGTATCAACTTCTCCAGACCATCACCT 628
Db 602 TGATCGACACCATCAAGGTCAAGGAGTGTGACCGGTATCAACTTCTCCAGACCATCACCT 661
Qy 629 TCGAGATCAACAAGATCCCAACACCGGTGTGATCACTCCGACTCCGACCAACATCAACA 688
Db 662 TCAAACTGACCGTGTGCTGACACTC-----ACCGGTACA 697
Qy 689 TGTGGATCGGTGACTTCTACATCTTCGCCAAGGAGTGGAGGTAAGGACATCAACATCC 748
Db 698 TGTGGATCAAAATCTTCAATCTGTTCGACAAAGACTGAAGGAAAGAAATCAAGACC 757
Qy 749 TGTTCAACTCCCTTGCGATACACCAACGTCGTCAGGACTACTGGGGTAAGACCTGAGAT 808
Db 758 TGTTCGACAAACCAAGTCCCAATCTGATCTCTGAAAGACTTCTGGGGTGAATCTGAGT 817
Qy 809 ACAACAAGGAGTACTACATGTTCAACATCGACTACTTGAACAGATACATGATCGCCAAC 867
Db 818 ACGACAACCGTACTACATGTTGATCTGTAGATCCGACAAATACGTTGACGTCAC 876

```

Search completed: November 7, 2002, 13:24:47  
 Job time : 3520 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 11:02:35 ; Search time 253 Seconds  
(without alignments)  
12203.517 Million cell updates/sec

Title: US-09-910-186A-9

Perfect score: 1371

Sequence: 1 gaattccgagatgaccatccc.....tctccgagtaaggaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1371	100.0	1371	21	AA287214
2	1371	100.0	1371	22	AA287214
3	665.2	48.5	3950	20	AA25521
4	649.8	47.4	1502	19	AAV30588
5	526.4	38.4	1208	21	AA298630
6	526.4	38.4	1440	21	AA298631
7	420.2	30.6	1374	22	AA254487
8	217.4	15.9	1347	21	AA287217
9	217.4	15.9	1368	22	AA254491

10	199.8	14.6	1317	22	AA254490	Botulinum toxin hea
11	198.2	14.5	1317	21	AA287216	DNA encoding synth
12	169.6	12.4	1351	19	AAV30576	Clostridium botuli
13	169.4	12.4	1323	22	AA254483	Botulinum toxin hea
14	169.4	12.4	1402	17	AAV29246	Type A neurotoxin
15	169.4	12.4	1402	19	AAV30572	Clostridium botuli
16	168.8	12.3	1326	22	AA254484	Botulinum toxin hea
17	168.8	12.3	1330	17	AAV29245	Type A neurotoxin
18	168.8	12.3	1330	19	AAV30571	Clostridium botuli
19	168.8	12.3	1332	22	AA254482	Botulinum toxin hea
20	168.8	12.3	1338	21	AA287212	DNA encoding synth
21	162.4	11.8	1400	21	AA287215	DNA encoding synth
22	162.4	11.8	1400	22	AA254488	Botulinum toxin hea
23	157.4	11.5	1278	22	AA254489	Botulinum toxin hea
24	151.8	11.1	1314	22	AA254499	Botulinum toxin hea
25	151.6	11.1	1313	18	AAV48101	Immunogenic type F
26	147.2	10.7	985	24	ABQ44816	Oligonucleotide fo
27	147.2	10.7	985	24	ABQ44817	Oligonucleotide fo
28	145.8	10.6	1332	22	AA254588	Sequence encoding
29	136	9.9	5163	19	AAV20700	Cryptosporidium pa
30	136	9.9	5163	21	AA261849	ORF encoding a por
31	136	9.9	5163	24	ABT04778	C parvum GP900 gen
32	136	9.9	5318	24	AA261848	DNA encoding a por
33	136	9.9	5318	24	ABT04777	C parvum GP900 gen
34	134.4	9.8	5318	19	AAV20701	Cryptosporidium pa
35	133.6	9.7	7417	23	ABT16368	Drosophila melanog
36	126.6	9.2	702	21	AA287221	DNA encoding BoNTA
37	122.8	9.0	1347	22	AA254589	Sequence encoding
38	121	8.8	2607	23	ABT16369	Drosophila melanog
39	120.4	8.8	1341	21	AA287213	DNA encoding synth
40	120.4	8.8	1341	22	AA254485	Botulinum toxin hea
41	115.4	8.4	5511	21	AA261847	Cryptosporidium pa
42	115.4	8.4	5511	24	ABT04776	C parvum GP900 gen
43	115.4	8.4	7334	21	AA261846	Cryptosporidium pa
44	115.4	8.4	7334	24	ABT04775	C parvum GP900 gen
45	110.6	8.1	3297	23	AA291437	DNA encoding novel

#### ALIGNMENTS

RESULT 1  
AA287214  
ID AA287214 standard; DNA; 1371 BP.  
XX  
AC AA287214;  
XX  
DT 08-MAY-2000 (first entry)  
XX  
DE DNA encoding synthetic BoNT serotype C (BoNTC) HC fragment.  
XX  
KW Botulinum neurotoxin; heavy chain; BoNT; serotype C;  
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
KW VEE; botulinum; vaccine; diagnosis; drug screening; ds.  
XX  
OS Clostridium botulinum.  
XX  
PH Key  
FT CDS  
FT Location/Qualifiers  
FT 10..1362  
FT /\*tag= a  
FT /product= "Synthetic botulinum neurotoxin serotype C  
(BoNTC) heavy chain C-terminal fragment (HC)"  
XX  
XX WO200002524-A2.  
XX  
XX  
PD 20-JAN-2000.  
XX  
XX  
PF 09-JUL-1999; 99WO-US15570.  
XX  
XX 10-JUL-1998; 98US-0092416.  
PR 12-MAY-1999; 99US-0133870.  
XX

(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;

WPI; 2000-160827/14.

P-PSDB; AAY77136.

Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum toxin serotypes A-G, is used for inducing an immune response against botulinum -

Disclosure: Page 41-42: 54pp: English:

The invention relates to novel vaccines that induce a protective immune response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant DNA construct comprising a vector, and at least one nucleic acid fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-G. In preferred embodiments of the invention, the vector is a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this vector results in the production of large amounts of a protein encoded by a sequence cloned into the replicon. The constructs are used to produce vaccines against botulism. The proteins can also be used as diagnostic tools for the diagnosis of botulism. The transformed host cells can be used to analyse the effectiveness of drugs and agents which inhibit toxin effects. The vaccine currently used against botulism is dangerous and expensive to produce, and contains formalin, which is very painful for the recipient. Also, the vaccine is incomplete, in that only 5 of the 7 serotypes are represented in the formulation. The novel vaccine of overcomes these problems, as it is easily purified, and available in large quantities. It is also expressed in the lymph nodes for a better immune response. Sequences AA287212-287217 represent synthetic DNA sequences encoding BoNT Hc fragments used in the present invention. These were optimised for codon usage for expression in yeast.

Sequence 1371 BP: 397 A: 416 C: 265 G: 293 T: 0 other:

```

Very Match      100.0%; Score 1371; DB 21; Length 1371;
t Local Similarity 100.0%; Pred. No. 4.1e-289;
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 GAATTACGATGACCATCCCATTCACACATCTTCTCTACACCAACTCCCCTGTTGAAG 60

1 GAATTACGATGACCATCCCATTCACACATCTTCTCTACACCAACTCCCTGTGAG 60

11 GACATCATCAACGAGTACTTCAACAACATCAACGACTCCAGATCCTGTCCCTGCAGAAC 120

1 GACATCATCAACGAGTACTTCAACAACATCAACGACTCCAAGATCCTGTCCCTGCAGAAC 120

1801 CGTAAGAACACCTTGGTCGACACCTCCGGTTACAACGCCGAGGTCTCCGAGGAGGGTGAC 1802

CGTAAGACACCTTGGTCGACACCTCCGGTTACAACGCCGAGGTCTCCGAGCAGGGTGAC 18

241 CTCCTGGTGAACCCCAATCTTCCCAATTCGACCTTCAAGCTGGGTTCTCCGGTGAGGACAGA 248

241 GTCCAGCTGACCCGACCTCTTCCCATTCGACCTTCAGCTGGTTCCTCCGGTGAGGACAGA 241

1 CCTGGCGTCTCCGCGCAGACCATCGTCATCGTCACTGACGAGTCCTC 30

[illegible][illegible][illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible][illegible]

QY	481	AACAAGCCTCCGGTTACAAAGTGGTCTTCGTCAACCATCAACATCATGCTGGT	540
Db	481	AACAAGCCTCCGGTTACAAAGTGGTCTTCGTCAACCATCAACATCATGCTGGT	540
QY	541	AACATGAAGATCTACATCAACGGTAAAGTATCGACACCAATCAAGGTCAAAGGATTCAC	600
Db	541	AACATGAAGATCTACATCAACGGTAAAGTATCGACACCAATCAAGGTCAAAGGATTCAC	600
QY	601	GGTATCAAGTCTCCAAAGACCATCACTTCGAGATCAACAAGATCCAGACACCGGTCG	660
Db	601	GGTATCAAGTCTCCAAAGACCATCACTTCGAGATCAACAAGATCCAGACACCGGTCG	660
QY	661	ATCACCTCCGACTCCGCAACATCAACATGTGGATCCGTGACTTCTACATTTCCGCAAG	720
Db	661	ATCACCTCCGACTCCGCAACATCAACATGTGGATCCGTGACTTCTACATTTCCGCAAG	720
QY	721	GAGTTGGCGGTAAAGACATCAACATCTCTGTTTCAACTCTTTGACGTACACCAAGCTCGTC	780
Db	721	GAGTTGGCGGTAAAGACATCAACATCTCTGTTTCAACTCTTTGACGTACACCAAGCTCGTC	780
QY	781	AAGGACTCTGGGTAACGACTGAGATACAAACAGGAGTACTACATGTTCAACATCGAC	840
Db	781	AAGGACTCTGGGTAACGACTGAGATACAAACAGGAGTACTACATGTTCAACATCGAC	840
QY	841	TACTTCAACAGATACATGTPACGCCAACTCCACACAGATGCTCTTCAACACACAGACGTAAC	900
Db	841	TACTTCAACAGATACATGTPACGCCAACTCCACACAGATGCTCTTCAACACACAGACGTAAC	900
QY	901	AACAACGACTTCAACGAGGGTTACAAAGATCATCATCAGCGTATTCAGAGGTAACACCAAC	960
Db	901	AACAACGACTTCAACGAGGGTTACAAAGATCATCATCAGCGTATTCAGAGGTAACACCAAC	960
QY	961	GACACCAAGTACAGAGGTGGTGACATCTCTGTTTCACTTCGACATGACTATCAACAAGGCC	1020
Db	961	GACACCAAGTACAGAGGTGGTGACATCTCTGTTTCACTTCGACATGACTATCAACAAGGCC	1020
QY	1021	TACAACCTGTTTATGAGAACGAGCCATGTACGCGGACACCTCCACCGAGGAGATC	1080
Db	1021	TACAACCTGTTTATGAGAACGAGCCATGTACGCGGACACCTCCACCGAGGAGATC	1080
QY	1081	TAGCCATCGGTTCGGTGAGCAGACCAAGGACATCAACGACACATCATCTTCAGATC	1140
Db	1081	TAGCCATCGGTTCGGTGAGCAGACCAAGGACATCAACGACACATCATCTTCAGATC	1140
QY	1141	CAGCCATGACACACTTACTACTACGGTTCCTCCAGATCTTCAAGTCCAACCTCAACGGT	1200
Db	1141	CAGCCATGACACACTTACTACTACGGTTCCTCCAGATCTTCAAGTCCAACCTCAACGGT	1200
QY	1201	GAGAACATCTCCGGTATCTGTTCCATCGGTACCTACAGATCCGCTCTGGGTGGTACTGG	1260
Db	1201	GAGAACATCTCCGGTATCTGTTCCATCGGTACCTACAGATCCGCTCTGGGTGGTACTGG	1260
QY	1261	TACAGACACAACTACTTGGTTCCAACTGTCAACAGGGTAACTACGCTCTCTTGGTGGAG	1320
Db	1261	TACAGACACAACTACTTGGTTCCAACTGTCAACAGGGTAACTACGCTCTCTTGGTGGAG	1320
QY	1321	TCCACTTCCACCACTGGGGATTCGTCCTCCAGTCTCCGAGTAATAGGAATC	1371
Db	1321	TCCACTTCCACCACTGGGGATTCGTCCTCCAGTCTCCGAGTAATAGGAATC	1371

## RESULTS

RESULT 2  
AAA54486

AAA54486  
ID AAA54486 standard; DNA; 1371 BP.

AA  
AC  
AAA54486:

11-APR-2001 (first entry)

Botulism toxin heavy chain C-terminal coding sequence (serotype C).

Botulism: toxin; neurotoxin; heavy chain; recombinant expression; XX

recombinant vector; antigen; immune response; vaccine; bacterium; KW

```

infection; ds.
XX
XX Synthetic.
OS Clostridium botulinum.
XX
XX Key Location/Qualifiers
XX CDS /*tag= a
FT /product= H_C peptide fragment
FT
XX
XX WO200067700-A2.
XX
XX 16-NOV-2000.
XX
XX 12-MAY-2000; 2000WO-US12890.
XX
XX 12-MAY-1999; 99US-0133865.
XX 12-MAY-1999; 99US-0133866.
XX 12-MAY-1999; 99US-0133867.
XX 12-MAY-1999; 99US-0133868.
XX 12-MAY-1999; 99US-0133869.
XX 12-MAY-1999; 99US-0133873.
XX 29-JUL-1999; 99US-0146192.
XX
XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
XX Smith LA, Byrne MP, Middlebrook JL, Iapenotiere H;
XX WPI; 2001-016048/02.
XX P-PSDB; AAB04092.
XX
XX New nucleic acids encoding the carboxy- or amino-terminal portions of
XX the heavy chain of botulinum neurotoxin of serotype A-G, useful as
XX vaccine against botulism
XX
XX Claim 2; Fig 5a; 73pp; English.
XX
XX Botulism neurotoxins are translated as a single 150 kDa polypeptide
XX chain and then posttranslationally nicked, forming a dichain
XX consisting of a 100 kDa heavy chain and a 50 kDa light chain which
XX remain linked by a disulfide bond. Nucleic acids encoding the
XX carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
XX chain of botulinum neurotoxin (BoNT) can be used in recombinant
XX expression vectors and expressed in transfected cells to produce
XX peptide antigens useful for eliciting an immune response to give
XX protective immunity against botulinum neurotoxin, which causes
XX botulism. The nucleic acids are expressible in a recombinant
XX organisms such as Escherichia coli or Pichia pastoris. The use
XX of recombinant nucleic acids are advantageous since it eliminates
XX the need to culture large quantities of hazardous toxin-producing
XX bacterium. Production yield from the genetically engineered product
XX is also high and cost of production is lower. The nucleic acids can
XX be derived from Clostridium botulinum serotypes A-G.
XX
XX Sequence 1371 BP; 397 A; 416 C; 265 G; 293 T; 0 other;
XX
XX Query Match 100.0%; Score 1371; DB 22; Length 1371;
XX Best Local Similarity 100.0%; Pred. No. 4.1e-289;
XX Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps
XX 0;
XX
XX 1 GAATTACGATGACATCCCATTCACATCTTCTCTACACCAACAACTCCCTGTTGAAG 60
XX |
XX 1 GAATTCAGGATGACATCCCATTCACATCTTCTCTACACCAACAACTCCCTGTTGAAG 60
XX |
XX 61 GACATCATCAACGAGTACTTCAACACATCAACGAGTCCCAAGATCCTGTCCTGCAGAAC 120
XX |
XX 61 GACATCATCAACGAGTACTTCAACACATCAACGAGTCCCAAGATCCTGTCCTGCAGAAC 120
XX |
XX 121 CGTAAGAACACTTGGTTCGACACTCCGGTTACACGCGGAGGTCTCCGAGGAGGGTGAC 180
XX |
XX 121 CGTAAGAACACTTGGTTCGACACTCCGGTTACACGCGGAGGTCTCCGAGGAGGGTGAC 180
XX |
XX 181 GTCCAGCTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTCTCTCCGCTGAGGACAGA 240
XX |

```

Db 1261 TACAGACACAACTACTTGGTTCCAACTGTCAAGCAGGGTAACCTACGCCCTCCCTGGTGGAG 1320  
Qy 1321 TCCACTTCCACCCACTGGGATTCGTCCTCCAGTCTCCGAGTAAATAGGAATTC 1371  
Db 1321 TCCACTTCCACCCACTGGGATTCGTCCTCCAGTCTCCGAGTAAATAGGAATTC 1371

RESULT 3  
AAAX25521  
ID AAAX25521 standard; DNA; 3950 BP.

XX AAAX25521;  
AC 02-AUG-1999 (first entry)  
XX DNA coding for modified botulinum toxin rBONT/C.  
DE Botulinum toxin; botulism; rBONT/C; vaccine; drug delivery;  
DE mutant; ss.

XX Clostridium botulinum.  
XX Synthetic.

XX Key Location/Qualifiers  
XX CDS 32..3907  
XX /\*tag= a

XX WO920306-A1.

XX 29-APR-1999.

XX 16-OCT-1998; 98WO-US21897.

XX 20-OCT-1997; 97US-0954302.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX Kiyatkin N, Maksymowych A, Simpson L;

XX WPI; 1999-302646/25.

XX P-PSDB; AA05814.

XX Modified toxin useful for systemic delivery of oral vaccines and  
XX therapeutic agents

XX Example 1; Page 26-31; 37pp; English.

This DNA sequence codes for a modified serotype C botulinum toxin, termed rBONT/C (see AA05814), in which amino acids His-229, Glu-230 and His-233 of the native sequence are substituted by Gly, Thr and Asn, respectively, i.e. the zinc binding motif (see AA05817) of the light chain holotoxin is modified, resulting in loss of endoprotease activity. DNA coding for the modified botulinum toxin was assembled from 3 separate toxin fragments using PCR and site-directed mutagenesis. The modified recombinant botulinum toxin maintains its ability to translocate from the gut into the general circulation but is non-toxic. It can be used as an oral vaccine for antigenic peptides, including botulinum toxin (i.e. an oral vaccine for botulism) or for the oral delivery of other therapeutic agents to the general circulation.

XX Sequence 3950 BP; 1589 A; 410 C; 616 G; 1335 T; 0 other;

Query Match 48.5%; Score 665.2; DB 20; Length 3950;  
Best Local Similarity 68.3%; Pred. No. 3e-135;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

Qy 13 ACCATCCCATTCACATCTCTCTCTACACCAACTCCCTGTTGAAGGACATCATCAAC 72  
Db 2558 ACAATACCTTTTAAATTTTTCATATACTAATAATCTTTTATAAAGATATAATTAAT 2617

Qy 73 GAGTACTTCAACACATCAACAGACTCCAGACTCCGCTCCGACAAAGCGTAAGACACC 132  
Db 73 GAGTACTTCAACACATCAACAGACTCCGCTCCGACAAAGCGTAAGACACC 132

Db 2618 GAATATTTCAATTAATTAATGATTCAAAAATTTTGGCCCTACAAAACAGAAAAAATACT 2677  
Qy 133 TTGCTCGACACCTCCCGGTTACAAAGCCGAGGTCTCCGAGAGGGTGAAGTCCAGCTGAC 192  
Db 2678 TTAGTGATCATCAGGATATAATGCAGAAGTGAAGAGAGGGGATGTTCCAGCTTAAT 2737  
Qy 193 CCAATCTCCCATTCGACTTCAAGCTGGGTTCCTCCGGTTCAGGACAGAGTAAAGTCAATC 252  
Db 2738 CCAATATTTCCATTTGACTTAAATAGGTAGTTTCAGGGGAGGATAGAGTAAGTTATA 2797  
Qy 253 GTCACCCAGAACAGAGAACATCGTCTCAACATCCATGTCAGAGTCCCTTCCTATCCTCTC 312  
Db 2798 GTAACCCAGAAATGAAAATATTTGATATATTTCTATGATGAAATTTTATAGCATTTAGT 2857  
Qy 313 TGGATCGAATCAACAAAGTGGTCTCCAACTTCGCCAGGTTCACACCATCATCGACTCCGTC 372  
Db 2858 TGGATTGAATAATAAATGAGTAAGTAATTTACCTGGATATATATTAATGATAGTGT 2917  
Qy 373 AAGAACAACTCCGTTGGTCCATCGGTATCATCTCCAACTTCCTCGATTCCTCCAGCTTC 432  
Db 2918 AAAAATAACTCAGGTGGAGTATAGGTATTTAGTAATTTTGTAGTATTTTACTTTTAAAA 2977  
Qy 433 CAGAACAGGACTCCGAGCAGTCCATCAACTCTCTCCAGCATCTCCCAACACGCTCCT 492  
Db 2978 CAAATGAAGATAGTGAACAAAGTAAATTTTAGTTATGATATATCAATAATGCTCCT 3037  
Qy 493 GGTACACAAAGTGTCTTCCTCCAGCTCACCACCAACATGATGGGTAACTGAAGATC 552  
Db 3038 GGATACAATAAATGGTTTTTGTAACTGTACTAACAATATGATGGGAAATATGAAGATT 3097  
Qy 553 TACATCAAGGTTAAGCTGATCGACACCATCAAGTCAAGGATTCACCGGTATCAACTTC 612  
Db 3098 TATATAAATGGAAAATTAATAGATACATATAAGTTAAAGTTAAAGTAAGTGAATATTT 3157  
Qy 613 TCCAAAGACCATCAGCTTCGAGATCAACAAGATCCAGACACCGGTCTGTATCACTCCGAC 672  
Db 3158 AGCAAACTATAACATTTGAATAAATAAATCCAGATACCGGTTTGTACTTACAT 3217  
Qy 673 TCCGACACATCAACAATGGTCCGCTGACTCTTACATCTTCGCCAGGAGTTGGACGGT 732  
Db 3218 TCTGATACATCAATATGTGGATAGAGATTTTATATATTTGCTAAAGAAATAGATGGT 3277  
Qy 733 AAGGACATCAACATCCTCTTCAACTCTTCAGTACACCACTCCGTCFCAAGGACTACTGG 792  
Db 3278 AAAGATATTAATATTTTAAATAGCTTGCAATATACTAATGTTGTAAGAAATTTATGG 3337  
Qy 793 GGTACGACCTGAGATACAACAAGAGTACTACATGTCACATCGACTACTTGAACAGA 852  
Db 3338 GGAAATGATTTAAGATATAATAAAGAAATTTATATGTTAATATAGATTTTAAATAGA 3397  
Qy 853 TACATGTACGCCAACTCCAGACAGATCTCTTCAACACACGACGTAAACAACGACTTC 912  
Db 3398 TATATGTATGCGAATCAGACAAATTTCTTTTAAACACGTAGAAATAATAATGACTTC 3457  
Qy 913 AACGAGGTTCAAGATCATCATCAAGCTATCAGAGGTACACCAACGACACGACATC 972  
Db 3458 AATGAGGATATAAATAATTAATAAAGAAATCAGAGAAATACAAATGATACATAGAGTA 3517  
Qy 973 AGAGGTGGTACATCCTGTACTTCGACATGACTATCAACAACAGGCTTCAACCTCTTC 1032  
Db 3518 CGAGGAGGAGATTTTATATTTTGTATGACAAATTAATAACAAGCATATAATTTGTT 3577  
Qy 1033 ATGAAGAAGACCATGTAGCCGACAAACACTCCAGGAGACATCTACGCCATCGGT 1092  
Db 3578 ATGAAGAATGAACATGTATGTCAGATAATCATAGTACTGAAGATATATATGCTATAGGT 3637  
Qy 1093 CTGCGTGACGACCAAGGACATCAACGACACATCATCTTCCAGATCCAGCCCAATGAAC 1152  
Db 3638 TTAAGAGAACCAACAAAGGATATAAATGATATATATTTCAATACACCAATGAAT 3697  
Qy 1153 AACACTTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAAGGTCAGAGACATCTCC 1212  
Db 3698 AATACITATTATTAGCATCTCAAAATTTAAATCAAAATTTTAAATGAGAGAAATATTCT 3757





Db 1195 ATCATAGTACTGAGATATATATGCTATAGGTTTTRAGAGACACAAACAGGATATAAATG 1254  
Oy 1121 ACACATCATCTTCCAGATCCAGCCAAATGAACAACTTACTACTACGTTCCCGAGATCT 1180  
Db 1255 ATAATATTATATTTCAATACAAACCAATGAATAATATTATTACGCATCTCAATAT 1314  
Oy 1181 TCAAGTCCAACTTCAAGGTTGAGAACATCTCCGGTATCTGTTCCATCGGTACTACTACAGT 1240  
Db 1315 TTAATCAATTTTAAATGAGAAATATTTCTGGAATATGTTCAATAGGTACTTATCGTT 1374  
Oy 1241 TCCGTCTGGGTGAGTACTGATCAGACACAACTACTTGTGTTCCACTGTCAAGCAGGTTA 1300  
Db 1375 TTAGACTTGGAGTGATTTGTTATAGACACAATTAATTTGGTGTGCTAGTGTGAAGCAAGAA 1434  
Oy 1301 ACTAGGCTCTGCTGCTGAGTCCAGCTTCCACCCAGTGGGATTCGTCCAGTCTCCGAGT 1360  
Db 1435 ATTATGCTTCATTAATAGAAATCAACTCACTCATTTGGGGTTTGTACCTGTAAAGTAAAT 1494  
1361 AATAG 1365  
1495 AAAAG 1499

RESULT 5  
AAZ98630  
AAZ98630 standard; DNA; 1208 BP.  
AAZ98630;  
20-JUN-2000 (first entry)  
C. botulinum type D toxin nucleotide sequence.

Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;  
protect; ds.

Clostridium botulinum.

WO200005252-A1.

03-FEB-2000.

20-JUL-1999; 99WO-IB01301.

22-JUL-1998; 98ZA-0006538.

(AGRI-) AGRIC RES COUNCIL.

De Bruyn EE, Botha AD;

WPI: 2000-205375/18.

P-PSDB; AAZ98982.

Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin  
type D, useful in vaccines for protection against botulism, comprises  
at least one amino acid mutation not present in the wild type D  
neurotoxins.

Claim 8; Page 54-57; 66pp; English.

This sequence represents a nucleotide sequence of a synthetic gene  
encoding a non-toxic immunogenic derivative of Clostridium botulinum type  
D toxin (BoNT). Botulinum neurotoxin causes botulism poisoning in cattle  
and sheep, and usually results in the death of the affected or poisoned  
animal. The non-toxic immunogenic fragments of the C. botulinum  
neurotoxin are useful in vaccines to protect animals (e.g. humans,  
cattle, sheep, pigs) against BoNT type D poisoning. The non-toxic  
fragments can be produced relatively simply and inexpensively  
(specifically by fermentation techniques). As the fragments are not  
toxic the risk to production staff is reduced.

Sequence 1208 BP; 416 A; 203 C; 216 G; 373 T; 0 other;

Query Match 38.4%; Score 526.4; DB 21; Length 1208;  
Best Local Similarity 65.9%; Pred. No. 3.7e-105;  
Matches 799; Conservative 0; Mismatches 401; Indels 12; Gaps 2;  
Oy 157 GCGAGGTCTCCGAGGAGGTGACGTCCAGCTGAACCCAAATCTTCCCAATTCGACTTCAAG 216  
Db 4 GCAGAAATTAGAGTTTGGGAAACGTGCAATGTAATCTTCTTCCATTTGACTTTAAG 63  
Oy 217 CTGGTTCCTCCCGTGGAGACAGAGTAAAGTTCATCGTCACCCAGAGAGAACATCGTC 276  
Db 64 TTGGTATGCTCTGTCATGACCGTGGCAAAATTAATGTAACCTCAAAAGAAATATGTT 123  
Oy 277 TAAACTCCATTCAGAGTCTTCTCCATCTCTCTCTGATCAGATCAACAGTGGGTC 336  
Db 124 TATAATGCTATGATGAGTCTTCTCTATGCTTCTGATCAGATTAATAATAGGTT 183  
Oy 337 TCAACTTCGCGGTTCACCATTCGACTCCGTCGTCAGAACAACTCCGGTTGGTCCATC 396  
Db 184 TCAATTTACAGGTTATACCATCATCGACGGTTAAATAATTCAGGATGCTTAT 243  
Oy 397 GGTATCATCTCCAACTTCTCTGCTTCCCTGATGACCGTGGCAAAATTAATGTAACCTCA 456  
Db 244 GGCATAATTAGCAATTTCTAGTTTTTACACTAAAGCAAAAGTAAAGTAGCGAGGAC 303  
Oy 457 ATCAACTTCTCTACGACATCTCCAAACGCTCTCTGGTTTACAAAGTGGTTCCTCGTC 516  
Db 304 ATTAATTTCTTATGACATTTCAAGAGCGCTGCTGGTTATAATAATGTTCTTCTGTA 363  
Oy 517 ACCGTCCACCAACATGATGGGTATACATGATGATCATCAACGGTAAAGTATCGAC 576  
Db 364 ACCATTACCAACATGATGGGTATGATGATGATGATGATGATGATGATGATGATGATG 423  
Oy 577 ACCATCAAGTCAAGGAGTGGAGCGGTATCACTTCTCCAAAGACCATCACTTCCGAGATC 636  
Db 424 ACCATCAAGTCAAGGAGTGGAGCGGTATCACTTCTCAAGACCATCACTTCCGAGATG 483  
Oy 637 AACAGATCCCGACACCGGTCTGATCCTCCGATCCGACATCAACATCAACATGTCGATC 696  
Db 484 AATAAAATCCCGACACAGGATTAATAACGTCGTGATCTGACAACTCAATATGTCGAT 543  
Oy 697 CGTGACTTCTACATCTTCCGACAGGAGTGGAGCGGTAAAGGACATCAACATCTCTGTTCAAC 756  
Db 544 AGATTTCTATATTTTCGCAAAAGATTAAGTACAGACAGACATTAATATCTTTTCAAT 603  
Oy 757 TCCTTGAGTACACCAACGTCGTCAAGGAGTACTGGGTTAACACCTGAGATACACAAAG 816  
Db 604 TCACITCAGTACACTAATGTTGTTAAGGATTAATGGGTTAATGATTAAGATATGATAAG 663  
Oy 817 GAGTACTACATGCTCACTACTTGAACAGATACATGTCAGCCCACTCCAGACAG 876  
Db 664 GAGTACTACATGATTAATGTAATTAATGATGATGATGATGATGATGATGATGATGAT 723  
Oy 877 ATCGTCTTCAACACGAGAGTAAACAAACGACTTCAACGAGGTTTACAAGATCATCATC 936  
Db 724 ATCGTCTTCAACACTCGTAAATAAACAAGATTTCAACGAGGTTTACAATAATATCATA 783  
Oy 937 AAGCGTATCAGAGTTAAACCAACGACAGAGTCAAGGTTGTCGATCCTGTCATCTC 996  
Db 784 AAAGAAATCCGTTGAAATACAAATGATTAAGTTCGTCGAGAACGTCGCTATATTT 843  
Oy 997 GACATGACTATCAACAAAGCGCTCAACCTGTTTCATGAAGACGAGACCATGTAGGCC 1056  
Db 844 AATAACAGATAGATAATAAGCAGTACTCTTTAGGCATGTATAAACCGTCAA----- 895  
Oy 1057 GACAACACTCCCGAGGAGCATCTACCCATCGTCTGCGTGGAGCAGACCAAGGACATC 1116  
Db 896 -GAAACTTAGGAGAGAGCTGTTCCATTTGGTGCATATAGATCAACGATGGAGGAAT 954  
Oy 1117 AACGACACATCATCTTCCAGATCCAGCAATGAACAACTTACTACTACGTTCCAG 1176  
Db 955 CGTAATATGGATCATTCATTTATCAACCATGTAATACGTTCCGATTACTACGCGAGCAA 1014  
Oy 1177 ATCTTCAAGTCCAACTTCAACGGTGAAGAACTCTCCGATATCTTCCATCGGTACCTAC 1236



QY 1294 CAGGCTACTAGCTCTCTCTGAGTCCACCTCCACCCAGGAGGATTCGCCAGTC 1353  
 Db 1189 ATCGAACATTAGCTTCTATTGATCACTTCTACACATGGGGCTTTGCCAGCG 1248  
 QY 1354 TCCGAGTAATAG 1365  
 Db 1249 TCTGAGTAAAG 1260

RESULT 7

AAAS4487  
 ID AAAS4487 standard; DNA; 1374 BP.  
 XX  
 AC  
 XX AAAS4487;

11-APR-2001 (first entry)

Botulinum toxin heavy chain C-terminal coding sequence (serotype D).

Botulinum toxin; neurotoxin; heavy chain; recombinant expression;  
 recombinant vector; antigen; immune response; vaccine; bacterium;  
 infection; ds.

Synthetic.  
 Clostridium botulinum.

Key Location/Qualifiers  
 CDS 10..1365  
 /tag= a  
 /product= H\_C peptide fragment

WO200067700-A2.

16-NOV-2000.

12-MAY-2000; 2000WO-US12890.

12-MAY-1999; 99US-0133865.  
 12-MAY-1999; 99US-0133866.  
 12-MAY-1999; 99US-0133867.  
 12-MAY-1999; 99US-0133868.  
 12-MAY-1999; 99US-0133869.  
 12-MAY-1999; 99US-0133873.  
 29-JUL-1999; 99US-0146192.

(USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

WPI; 2001-016048/02.  
 P-PSDB; AAB04093.

New nucleic acids encoding the carboxy- or amino-terminal portions of  
 the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 vaccine against botulism

Claim 2; Fig 6a; 73pp; English.

Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
 chain and then posttranslationally nicked, forming a dichain  
 consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 remain linked by a disulfide bond. Nucleic acids encoding the  
 carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 expression vectors and expressed in transformed cells to produce  
 peptide antigens useful for eliciting an immune response to give  
 protective immunity against botulinum neurotoxin, which causes  
 botulism. The nucleic acids are expressible in a recombinant  
 organisms such as *Escherichia coli* or *Pichia pastoris*. The use  
 of recombinant nucleic acids are advantageous since it eliminates  
 the need to culture large quantities of hazardous toxin-producing  
 bacterium. Production yield from the genetically engineered product

CC is also high and cost of production is lower. The nucleic acids can  
 CC be derived from Clostridium botulinum serotypes A-G.

XX Sequence 1374 BP; 367 A; 423 C; 274 G; 310 T; 0 other;

Query Match 30.6%; Score 420.2; DB 22; Length 1374;  
 Best Local Similarity 63.0%; Pred. NO. 5.1e-82;  
 Matches 754; Conservative 0; Mismatches 398; Indels 45; Gaps 5;

QY 4 TTCACGATGACCATCCCATTCACATCTCTCTCTACCAACAACCTCCCTGTTGAAGGAC 63  
 Db 40 TGCAGAACACCATGCCATTCACATCTCTCTCTACCAACAACCTCCCTGTTGAAGGAC 99  
 QY 64 ATCATCAACGAGTACTTTCACAAACATCAACGACTCCAAGATCTGTCCTGCAAGACCT 123  
 Db 100 ATCATCAACGAGTACTTTCACATCCATCAACGACTCCAAGATCTGTCCTGCAAGACAA 159  
 QY 124 AAGAACACCTTGGTCGACACTCCGGTTACACGCGAGGCTCTCCGAGGAGGTGACGTC 183  
 Db 160 AAGAACGCTTGGTCGACACTCCGGTTACACGCGAGGCTGAGTGGTGACACGTC 219  
 QY 184 CAGCTGAACCCCAATCTTCCCATTCGACTTCAAGCTGGGTTCTCCGCTGAGGACAGGT 243  
 Db 220 CAGTTGAACACCATCTACACCAAGCACTTCAAGTTGTCTCTCTCCGGTGA-----C 270  
 QY 244 AAGGTCACTGTCACCGAGAGAGACATCGTCTACAACCTCCATGTACGACTCTCTCC 303  
 Db 271 AAGATCATCTGTCACCTTGAACACACATCTTGTCTCGGCACTTACGAGAACTCTCT 330  
 QY 304 ATCTCCTTCTGGATCAGAAATCAACAGTGGGTCTCCAACCTTGC---CAGGTTACACCATC 360  
 Db 331 GTCCTCTCTGGATCAAGATCTCCAAAGGACTTGACCAACTCCCAACAGAGTACACCATC 390  
 QY 361 ATCGACTCCGTCAGAGAACACTCCGGTTGGTCCATCGGTATGATCTCCACTTCTCTGGTC 420  
 Db 391 ATCAACTCCATCGAGAGAACTCCGGTTGGAGTTGTGTATCGGTAACTGACGTACATCGAG 450  
 QY 421 TTCACCTGGAAGCAGAGGAGGACTCCGAGCAGTCCATCAACTCTCTCCACGACATCTCC 480  
 Db 451 TGGATCTTGCAGAGGAGCTCAACCGTAAGTCAAGTCTCTGATCTTCGACTACTCCGAGTCC 510  
 QY 481 AACAAAGCTCTCTGGTTA---CAACAAGTGGTTCTTGTACCGTCCACCAACACATGATG 537  
 Db 511 TTGTCCACACACCGGTTACACCAACAGTGGTTCTTGTACCGTCCACCAACACATGATG 570  
 QY 538 GGTACATGAAGATCTACATCAACGGTAAAGTGTATCGATCGACACCATCAAGGTCAAGGAGTTG 597  
 Db 571 GGTACATGAAGTGTATCAATCAACGGTAAAGTGTATCGATCGACACCATCAAGGTCAAGGAGTTG 630  
 QY 598 ACCGGTATCAACTTCTCCAGACCATCACTTCCGATGATCAAGAGATCCAGACACCGGT 657  
 Db 631 GACGAGGTCAAGCTGGACAAAGACCATCTCTCGGTATCGATCGACACCATCAAGTCAAGTCA 683  
 QY 658 CTGATCACCTCCGACTCCGCAACATCAACATGTGGATCGGTGACTTCTACATCTTCTCGCC 717  
 Db 684 -----CGAGAACCAAGTGTGTGGATCGGTGACTTCTACATCTTCTCTCC 726  
 QY 718 AAGGAGTTGGAGGATAGGACATCAACATCTCTGTCAACTCTCTGAGTACACCAACGTC 777  
 Db 727 AAGGAGTGTCCAACAGGAGACATCAACATCTCTGATCGAGGTCAGATCTCTGAGGAGGTC 786  
 QY 778 GTCAGGAGTACTGGGTAGGACCTGAGATACAAAGGAGGAGTACTACATGTTCAACATC 837  
 Db 787 ATCAAGGAGTACTGGGTAAAGTCTGAGTCAAGGAGGAGTACTACATGTTCAACATC 846  
 QY 838 GACACTTGAACAGATACATGATAGCCAACTCCAGACAGATCGTCTTCAACACCAAGACGT 897  
 Db 847 AACTACATCGACCGTTATACGCCCCAGAGTCCCAAGTCTCTGCTGGTCCAGTACCCT 906  
 QY 898 AACAAACAGGATTCACAGGAGGTTACAAAGATCATCATCAAGGATATCAGAGGTACAC 957  
 Db 907 GACCGTTCCAAAGCTGTACACCGGTAACCCCTATCACCATCAAGTCCGCTCTCCGCAAGAAC 966

QY 958 AACGACACGAGCTGAGGCTGGTACATCTCTGCTACTTCGACATGACATCAACAGCAAG 1017  
 Db 967 CTTTACTTCCCTGATCTGACGCTGACATCACTCTCTGACATGCTGTACACACTCCCGT 1026  
 QY 1018 GCTTACAACTGTCATGATGAGAACGACGACATGATGAGGCAAA-----CCACTCCACC 1071  
 Db 1027 AAGTACATGATCATCGCTGACACCGACACCATCTAGCGCACCCAGGGTGGTGTCTCC 1086  
 QY 1072 GAGGACATCTACCCATCGCTGCTGAGCAGACCAAGGACATCAAGGACACATCATC 1131  
 Db 1087 CAGAACTGTCTACGCCCTGAAAGCTGCACTCCACCTGGGTAATGATCGGTATC 1146  
 QY 1132 TTCAGATCAGCAATGACACACATCTACTACTAGCTTCCAGATCTTCAAGTCC 1188  
 Db 1147 TTCTCCATCAAGACATCGTCTCCAGAACAGTACTGCTCCAGATCTTCTCCCTC 1203  
 RESULT 8  
 AA287217  
 AA287217 standard; DNA; 1347 BP.  
 AA287217;  
 08-MAY-2000 (first entry)  
 DNA encoding synthetic BoNT serotype G (BoNTG) Hc fragment.  
 Botulinum neurotoxin; heavy chain; BoNT; serotype G;  
 C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
 VEE; botulinum; vaccine; diagnosis; drug screening; ds.  
 Clostridium botulinum.  
 Synthetic.  
 Key Location/Qualifiers  
 CDS 1..1347  
 /tag= a  
 /product= "Synthetic botulinum neurotoxin serotype G  
 (BoNTG) heavy chain C-terminal fragment (Hc)"  
 /note= "No stop codon given in the specification"  
 WO200002524-A2.  
 20-JAN-2000.  
 09-JUL-1999; 99WO-US15570.  
 10-JUL-1998; 98US-0092416.  
 12-MAY-1999; 99US-0133870.  
 (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
 Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;  
 WPI; 2000-160827/14.  
 P-PSDB; AAY77139.  
 Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
 toxin serotypes A-G, is used for inducing an immune response against  
 botulinum -  
 Disclosure; Page 47-48; 54pp; English.  
 The invention relates to novel vaccines that induce a protective immune  
 response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
 and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
 DNA construct comprising a C-terminal heavy chain fragment (Hc) from BoNT  
 fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT  
 serotypes A-G. In preferred embodiments of the invention, the vector is  
 a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of  
 this vector results in the production of large amounts of a protein  
 encoded by a sequence cloned into the replicon. The constructs are used  
 to produce vaccines against botulinum. The proteins can also be used as

CC diagnostic tools for the diagnosis of botulism. The transformed host  
 cells can be used to analyse the effectiveness of drugs and agents which  
 inhibit toxin effects. The vaccine currently used against botulism is  
 dangerous and expensive to produce, and contains formalin, which is very  
 painful for the recipient. Also, the vaccine is incomplete, in that only  
 5 of the 7 serotypes are represented in the formulation. The novel  
 vaccine overcomes these problems, as it is easily purified, and  
 available in large quantities. It is also expressed in the lymph nodes  
 for a better immune response. Sequences AA287212-287217 represent  
 synthetic DNA sequences encoding BoNT Hc fragments used in the present  
 CC. invention. These were optimised for codon usage for expression in yeast.  
 xx  
 SQ Sequence 1347 BP; 358 A; 442 C; 263 G; 284 T; 0 other;  
 Query Match 15.9%; Score 217.4; DB 21; Length 1347;  
 Best Local Similarity 52.5%; Pred. No. 7.6e-38;  
 Matches 631; Conservative 0; Mismatches 516; Indels 54; Gaps 5;  
 QY 43 AACAATCCCTGTTGAAGGACATATCAACGAGTACTTCAACAACATCAAGACTTCCAA 102  
 Db 4 AAGGACACCATCTGATCCAGGCTTCTCAACAACATCAATCTCCACATCTCTCCAAACGCC 63  
 QY 103 ATCTGTCTCTGACAGAACCGTAAGAACAACCTTGTGCGACACCTCGGTTAGAACGCCGAG 162  
 Db 64 ATCTGTCTCTGACAGAACCGTAAGAACAACCTTGTGCGACACCTCGGTTAGAACGCCGAG 123  
 QY 163 GTCTCGAGGAGGTTGACGTCCAGCTGAACCAATCTTCCCAATTCGACTTCAAGTGGGT 222  
 Db 124 ATCAAGCTGCTTCCGACGCTATCTTCAACGACATCGGTAAAGGTCAGTTCAGGTGAAC 183  
 QY 223 TCCTCCGGTGAGCAGAGAGTGAAGTCACTGTCACCGCCAGAGAGAAATCGTCTACAC 282  
 Db 184 AACTCC-----GAGAACTCCACATCACCGCCACAGTCCCAAGTTCGTCTACAC 237  
 QY 283 TCATGTACGAGTCTTCTTCCATCTCTTGTGATCAGAAATCAACAAGTGGTGTCTCCAA 342  
 Db 238 TCATGTGACAACTTCTCCATCAACTTCTGGTCCGTACCCCAAGTACAAACAAC 297  
 QY 343 -----TTGCCAGTTACACCATCATCTGACTCCGCTGATCAAGAACACTCC 384  
 Db 298 GACATCCAGACCTACCTGCGAGAACGAGTACACCATCATCTCTGTATCAAGAACACTCC 357  
 QY 385 GGTTGGTCCATCGGTATCATCTCCAACTTCTCCCTGGTCTTCCACCTGAAGCAGAACGAGAC 444  
 Db 358 GGTTGGAGGTCTCCATCAAGGAAACCGTATCATCTGGACCCCTGATCGAGCTCAACGCC 417  
 QY 445 TCCGAGCAGTCCATCAACTTCTCTAGGACATCTCCAGACGCTCTCTGGTTA---CAAC 501  
 Db 418 AAGTCCAAAGTCCATCTTCTCGAGTACTCCATCAAGGACACATCTCCGACTACATCAAC 477  
 QY 502 AAGTGGTCTTCTGTCACCGTCAACCAACATGATGGGTAAACATGAAGATCTACATCAAC 561  
 Db 478 AAGTGGTCTTCCATCACCATCAACCAAGCCGCTGAGTGAAGCCACATCTACATCAAC 537  
 QY 562 GGTAAGTGTATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCCAAAGACC 621  
 Db 538 GGTTCCCTGAGAAAGTCCGAGAAAGATCTCAACTGGACCGTATCAACTTCTCCAAAGACC 597  
 QY 622 ATCACTTCCAGATCAACAGAGATCCCGACACCGGCTCTGATCACTCCGACTCCGACAC 681  
 Db 598 ATCACTTCAAG-----CTGATCACTGTACCCACACACC 633  
 QY 682 ATCAACATGTGGATCCGTGACTTCTACATCTTGGCCAAAGGAGTGTGACGGTAAAGCATC 741  
 Db 634 AAGTTCGTCTGGATCAAGGACTTCAACATCTTGGTCCGTGAGCTGAACCCACCGAGGTC 693  
 QY 742 AACATCTGTCTCACTCTTGTGAGTACACCAAGTGTCTCAAGGACTTACGGGTACGAC 801  
 Db 694 TCCCTCCCTGACTGATTCAGTCTCTCCACACACCGCTGAAGGACTTCTGGGGAACCCA 753  
 QY 802 CTGAGATCAACAAAGGAGTACTACATGGGTCAACATCGACTACTTGAACAGATACATGTAC 861  
 Db 754 CTGGTTACGACCCAGTACTACCTGTTCAACCCAGGGTATGACAGAACATCTACATCAAG 813

QY 862 GCCAACTCCA---GACAGATCGTCTTCAACACGAGAGGTAAACAAACAGTCTCAACGAG 918  
 DB 814 TACTTCTCCAAAGGCTCCATGGGTGAGACCGCCCTCGTACCAACTTCAACAGCGCGC 873  
 QY 919 GGTAAAGATCATCATCAAGCGTATCAGAGGTAAACACCAACGACACGAGTCAAGGT 978  
 DB 874 ATCAACTACCAAGACCTGTACTCTGGGTCTGGTTTCATCATCAAGAAGGCTCCAACTCC 933  
 QY 979 GGTGATCTCTGTACTTGCATCATGATCATCAACAAAGGCTTCAACCTGTTCATGAAG 1038  
 DB 934 CGTAACATCAACAAAGCAACATCGTCGGTGGGTGACTTACATCTACCTGAACATCGAC 993  
 QY 1039 AAGGAGACATGTACGCGCAACCACTTCCACGAGGACATCTACGCCATCGGCTCTGCGT 1098  
 DB 994 RACATCTCGAGGAGTCTACCGTGTCTAGTCTCTGGTCACTCCAGGAGATCCAGACC 1053  
 1099 GAGCAGACAGGACATCAACGACAAATCATCTTCCAGATCCAGCATGACATGACACACT 1158  
 1054 CAGCTGTCTCTGGGCCCCAATCAACGAGGACCCCTACCTTCTACGACGCTCTGCGATCAAG 1113  
 1159 TACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGAACATCTCCGGTATC 1218  
 1114 RAGTACTAGGAGACCACTTACATCTGTCAGATCTCTGCGAGAGGACCAAGACC 1173  
 1219 T 1219  
 1174 T 1174  
 CDS  
 AAAS4491  
 AAAS4491 standard; DNA; 1368 BP.  
 AAAS4491;  
 11-APR-2001 (first entry)  
 Botulinum toxin heavy chain C-terminal coding sequence (serotype G).  
 Botulinum toxin; neurotoxin; heavy chain; recombinant expression;  
 recombinant vector; antigen: immune response; vaccine; bacterium;  
 infection; ds.  
 Synthetic.  
 Clostridium botulinum.  
 Key Location/Qualifiers  
 CDS 10..1359  
 /\*tag= a  
 /product= H\_C peptide fragment  
 WO200067700-A2.  
 16-NOV-2000.  
 12-MAY-2000; 2000WO-US12890.  
 PF 12-MAY-1999; 99US-0133865.  
 PR 12-MAY-1999; 99US-0133866.  
 PR 12-MAY-1999; 99US-0133867.  
 PR 12-MAY-1999; 99US-0133868.  
 PR 12-MAY-1999; 99US-0133869.  
 PR 12-MAY-1999; 99US-0133873.  
 PR 29-JUL-1999; 99US-0146192.  
 (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
 Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
 WPI; 2001-016048/02.  
 DR P-PSDB; AAB04167.  
 XX

PT New nucleic acids encoding the carboxy- or amino-terminal portions of  
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 XX vaccine against botulism  
 FS Claim 2; Fig 10a; 73pp; English.  
 XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
 CC chain and then posttranslationally nicked, forming a dichain  
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 CC remain linked by a disulfide bond. Nucleic acids encoding the  
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 CC expression vectors and expressed in transformed cells to produce  
 CC peptide antigens useful for eliciting an immune response to give  
 CC protective immunity against botulinum neurotoxin, which causes  
 CC botulism. The nucleic acids are expressible in a recombinant  
 CC organisms such as Escherichia coli or Pichia pastoris. The use  
 CC of recombinant nucleic acids are advantageous since it eliminates  
 CC the need to culture large quantities of hazardous toxin-producing  
 CC bacterium. Production yield from the genetically engineered product  
 CC is also high and cost of production is lower. The nucleic acids can  
 CC be derived from Clostridium botulinum serotypes A-G.  
 XX  
 SQ Sequence 1368 BP; 366 A; 445 C; 267 G; 290 T; 0 other;  
 Query Match 15.9%; Score 217.4; DB 22; Length 1368;  
 Best Local Similarity 52.5%; Pred. No. 7.6e-38;  
 Matches 631; Conservative 0; Mismatches 516; Indels 54; Gaps 5;  
 QY 43 AACAACTCCCTGTTGAAGGACATCATCAAGAGTACTTCAACAACTCAACGACTCCAAAG 102  
 DB 13 AAGGACACCATCTGATCCAGGCTTTCACAACTACATCTCCAAATCTCTCTCAACGCC 72  
 QY 103 ATCCTGTCCCTGACAGACCGTAAAGAACACCTTGGTGGACACCTCCGGTTACACGCCGAG 162  
 DB 73 ATCCTGTCCCTGCTCCTACCGGTGGTGGTCTGATCGACTCTCCGGTTACGGAGCACCC 132  
 QY 163 GTCTCCGAGGAGGTGCTCCAGCTGAAACCAATCTTCCATTCCGACTTCAAGCTGGGT 222  
 DB 133 ATGACGTCGGTTCGAGGTCATCTTCAACGACATCGTAAACGTCAGTCAAGCTGAAC 192  
 QY 223 TCTCCGCTGAGGACAGAGGTAAGGTATCGTCAACCAAGAACGAGACATCGTCTCAAC 282  
 DB 193 AACTCC-----GAGAACTCCAACTACACGCCCCACACGTCGAGTCTGCTACGAC 246  
 QY 283 TCCATGTACGAGTCTCTCCATCTCTTGGATCAGATCAACAAAGTGGGTCTCCAAAC 342  
 DB 247 TCCATGTTCGAACTTCTCCATCACTTTCGGTCCGTACCCCAAGTACACACACAC 306  
 QY 343 -----TTGCCAGGTTACACCATATCGACTCCGCTCAAGAACACTCC 384  
 DB 307 GACATCCAGACCTACCTGCAGAACGAGTACACCATCATCTCTCTGTATCAAGAACGACTCC 366  
 QY 385 GGTGGTCCATCGGTATCATCTCCAACTTCTGGTCTTCCCTGAGGAGACGAGGAC 444  
 DB 367 GGTGGAGAGGTCCTCATCAAGGGAACCGTATCATCTGTGGACCTGTGTCGACGTCAACGCC 426  
 QY 445 TCCGAGCAGTCCATCACTTCTCTAGCATCTTCCAAACACGCTCTCTGTTA---CAAC 501  
 DB 427 AAGTCCAAAGTCCATCTTCTCGAGTACTCCATCAAGGACAACTCTCCGACTACATCAAC 486  
 QY 502 AAGTGGTCTTTCGTCACCGTCAACCAACATGATGGTGAACATGAGATCTACATCAAC 561  
 DB 487 AAGTGGTCTTTCATCATCATCAACGAGCGCTGTGGTGAACGCAACATCTACATCAAC 546  
 QY 562 GGTAAAGTGTGACGACCATCAAGGTCAAGGAGTGTGACCGGTATCAACTTCTCCAAAGACC 621  
 DB 547 GGTTCCTTGAAGAGTCCGAGAGATCTCTGAACCTGGACCGGTATCAACTCTCTCAACGAC 606  
 QY 622 ATCACTTTCGAGATCAACAAAGATCCAGACACCGGTCTGTATCACTCCGACTCCGACAC 681  
 DB 607 ATCGACTTCAAG-----CTGATCAACTGTACCGACACCAACC 642



480 CTACATCAACAGTGGATCTTCGTCACCATCAACATCAACCGTTTGGGAACCTCCAGAT 539  
 552 CTACATCAACGGTAAAGTGTGCGACCAATCAAGGTCAGGAGTTGACCGGTATCAACTT 611  
 540 CTACATCAACGGTAAAGTGTGCGACCAATCAAGGTCAGGAGTTGACCGGTATCAACTT 599  
 612 CTCGAAGACCATCACTTCGAGATCAACAGATCCCGAGACACCGGTCTGTATCACCCTCGA 671  
 600 CTCGACCAACATTTGTCAAGATCGTGGTTTAAGACGACCGGT----- 646  
 672 CTCGACCAACATCACTTCGAGTTCGTCATCTTACATCTTCGCAAGGAGTTGGACGG 731  
 647 -----ACGTCGGGATCCGTTACTTCAAAAGTCTCGACACTGAGTTGGTAA 692  
 732 TAAGGACATCACTCTGTTCAACTCTTCGAGTACACCAAGTGTGTCAGGACTACTG 791  
 693 GACCGAGATCGAGACCTTGTACTCCGAGCGCTGACCCATCCATCCTGAAGGACTTCG 752  
 792 GGGTAACGACCTGAGATACACAAAGGAGTACTACATGGTCAACATCGACTACTTGAACAG 851  
 753 GGGTAACCTACCTGCTGTACACAAAGCTTACTACTTCTGCTGAACCTTGTGGTACCGACAA 812  
 852 ATAGATGACGCCAACTCCAGACAGATCGTCTCAAC 888  
 813 GTCCATCACCAGAACTCCAACTCTTTGAACATCAAC 849  
 RESULT 11  
 287216  
 AA87216 standard; DNA; 1317 BP.  
 AA87216;  
 08-MAY-2000 (first entry)  
 DNA encoding synthetic BoNT serotype F (BoNTF) Hc fragment.  
 Botulinum neurotoxin; heavy chain; BoNT; serotype F;  
 C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
 VEE; botulism; vaccine; diagnosis; drug screening; ds.  
 Clostridium botulinum.  
 Synthetic.  
 Key Location/Qualifiers  
 CDS 10..1308  
 /tag= a  
 /product= "Synthetic botulinum neurotoxin serotype F  
 (BoNTF) heavy chain C-terminal fragment (Hc)"  
 WO200002524-A2.  
 20-JAN-2000.  
 09-JUL-1999; 99WO-US15570.  
 10-JUL-1998; 98US-0092416.  
 12-MAY-1999; 99US-0133870.  
 (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
 Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;  
 WPI; 2000-160827/14.  
 P-PSDB; AAY7138.  
 Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
 toxin serotypes A-G, is used for inducing an immune response against  
 botulinum -  
 Disclosure; Page 45-46; 54pp; English.

The invention relates to novel vaccines that induce a protective immune  
 response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
 and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
 DNA construct comprising a C-terminal heavy chain fragment (Hc) from BoNT  
 serotypes A-G. In preferred embodiments of the invention, the vector is  
 a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of  
 this vector results in the production of large amounts of a protein  
 encoded by a sequence cloned into the replicon. The constructs are used  
 to produce vaccines against botulism. The proteins can also be used as  
 diagnostic tools for the diagnosis of botulism. The transformed host  
 cells can be used to analyse the effectiveness of drugs and agents which  
 inhibit toxin effects. The vaccine currently used against botulism is  
 dangerous and expensive to produce, and contains formalin, which is very  
 painful for the recipient. Also, the vaccine is incomplete, in that only  
 5 of the 7 serotypes are represented in the formulation. The novel  
 vaccine of overcomes these problems, as it is easily purified, and  
 available in large quantities. It is also expressed in the lymph nodes  
 for a better immune response. Sequences AA87212-287217 represent  
 CC synthetic DNA sequences encoding BoNT Hc fragments used in the present  
 CC invention. These were optimised for codon usage for expression in yeast.  
 XX  
 SQ Sequence 1317 BP; 397 A; 375 C; 239 G; 305 T; 1 other;  
 Query Match 14.5%; Score 198.2; DB 21; Length 1317;  
 Best Local Similarity 56.6%; Pred. No. 1.1e-33;  
 Matches 496; Conservative 0; Mismatches 333; Indels 48; Gaps 5;  
 QY 27 CATCTTCTCTACACCAACCTCCCTGTTGAAGGACATCATCAACGAGTACTTCAACAA 86  
 DB 6 CACGATGTCCTACCAACGACGACAGATCTTGATCTGTACTTCAACAGCTGTACAAGA 65  
 QY 87 CATCAACGACTCCAAAGATCTCTGTCGAGACCGGTAGAACACCTTGGTCGACACCTC 146  
 DB 66 GATCAAGGACAACTCCATCTTGGACATGAGATACGAAACAAATAGTTTCATCGACATC 125  
 QY 147 CGGTTACAAACGCGAGGTCTCCGAGGAGGTGACGTCAGCTGAACCCATCTCCCAT 206  
 DB 126 CGGTTAGGTTCCAAACATCTCCATCAACGCTGACGTCTACATCTATCCACCAATAGA 185  
 QY 207 CGATTCAAGTCGGTTCCTCGGTGAGGACAGAGTATGATCGTCAACCGACGACGA 266  
 DB 186 CCAGTTCGGATCTACTCTCTCC-----AAGCCCTCCGAGGTCAACATCGCTCAGAAC 239  
 QY 267 GAACATGCTCTACAACTCCATGTACGAGTCTCTCTCATCTCTCTGATCATCAGATCAA 326  
 DB 240 CGACATATCTACAAACGAGATACACGACACTTCTCATCTCTCTGTCGATATCC 299  
 QY 327 CAAGTGGTCTCCAACTT-----GCCAGGTTACACCATCATCGACTCCCTC---AA 374  
 DB 300 AAAGTACTTCAACAAAGGTCAACCTGAAATACGAGTACACCATCATCGACTCCGTAA 359  
 QY 375 GAACAACTCCGGTGTCTCCATCGGTATCTCCAACTTCTCTGTCCTTCAACCTGAAGCA 434  
 DB 360 CAATACTCCGGATGGAAGATCTCCCTGACTACACAGATCATCTGTGGACCTCGAGGA 419  
 QY 435 GAACGAGACTCCGAGCAGTCCATCAACTTCTCTCTAG---GACATCTCCAAACACCTCC 491  
 DB 420 CACGCGCGGTAAACAATCAGAAGTTGGTCTTCACTACACNAGATGATCTCCATCTCCGA 479  
 QY 492 TGGTTACAAACAGTGGTCTCTCGTCACCGTCAACCAACATCATGGTTAATCATGAGAT 551  
 DB 480 CTACATCAACAAAGTGGATCTTCGTCACCATCAACCAATACCGTTTGGGAACCTCCAGAT 539  
 QY 552 CTACATCAACGAGTGTGATCGACACATCAAGGTCAGGAGTTGACCGGTATCAACTT 611  
 DB 540 CTACATCAACGAGTGTGATCGACGAGAGTCCATCTCCAACTTGGGTGACATCCACGT 599  
 QY 612 CTCGAAGACCATCACTTCGAGATCAACAGATCCCGAGACACCGGTCTGTATCACCCTCGA 671  
 DB 600 CTCGACCAACATTTGTTCAGATCGTGGTGTGAAGACACCGCTT----- 646  
 QY 672 CTCGACCAACATCAACATGTGGATCCGTCGACTTCTACATCTTCGCAAGGAGTTGGACGG 731



```

Db 647 -----ACCTGGGATCCGTTACTTCAAGTCTTCGACACTGAGTTGGTAA 692
Qy 732 TAAGGACATCAACATCTGTTCAACTCCCTTGCAGTACACCAACGCTGCTCAAGGACTACTG 791
Db 693 GACCGAGATCGAGACCTTGTTACTCCGACGAGCTGACCCATCCATCCTGAAGGACTTCTG 752
Qy 792 GGTAAAGACCTGAGATACAAAGGNGTACTACATGCTCAACATCGACTACTTGAACAG 851
Db 753 GGTAAATACCTGCTGTACAAACAAAGTTACTACTTCTGTAACCTGTTGCGTACCGACAA 812
Qy 852 ATACATGACGCCAACCCACAGACAGATCGTCTTCAAC 888
Db 813 GTCCATCAACCCAGAACTCCAACTCTTGAACATCAAC 849

RESULT 12
AAV30576
AAV30576 standard; DNA; 1351 BP.
AAV30576;
07-DEC-1998 (first entry)
Clostridium botulinum toxin A fragment C gene in pHisBotA(syn).
Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
botulism; ds.
Clostridium botulinum serotype A.
Synthetic.
Key Location/Qualifiers
CDS 1..1338
/*tag= a
W09808540-A1.
05-MAR-1998.
28-AUG-1997; 97WO-US15394.
28-AUG-1996; 96US-0704159.
(OPHI-) OPHIDIAN PHARM INC.
Thalley BS, Williams JA;
WPI; 1998-230234/20.
P-PSDB; AAW68391.
Host cell containing recombinant expression vector encoding
Clostridium botulinum type B or E toxin - useful to treat humans
and other animals at risk of intoxication with clostridial toxin
Example 29; Page 279-281; 428pp; English.
This is the DNA sequence of the Clostridium botulinum serotype A
toxin C fragment gene contained in plasmid pHisBotA(syn). The
encoded toxin A polypeptide (see AAW68391) has a histidine-tagged
N-terminal extension. The vector was used to express native
(i.e. non-fusion) soluble C fragment in Escherichia coli host
cells. The invention relates to recombinant proteins derived from
C. botulinum toxins. Methods are provided which allow for the
isolation of soluble recombinant proteins free of significant
endotoxin contamination. Preferred hosts for production of
recombinant proteins are E. coli, insect cells and yeast cells.
The recombinant toxins are used as immunogens for the production
of vaccines and antitoxins that are useful in the treatment of
humans and animals at risk of intoxication with clostridial toxin.
Sequence 1351 BP; 407 A; 348 C; 247 G; 349 T; 0 other;

```

```

Query Match 12.4%; Score 169.6; DB 19; Length 1351;
Best Local Similarity 54.4%; Pred. No. 28-27; Indels 42; Gaps 4;
Matches 455; Conservative 0; Mismatches 339;
Qy 44 ACAACTCCCTGTTGAAGGACATCATCAACGAGTACTTCAACAACTCAACGACTCCAAGA 103
Db 20 ACAUGGCTCGTCTGCTTACCTTACATGANTACATCAAGAACATCATCAATCACTCCA 79
Qy 104 TCCTGTCCTCGAGAACCGTAAAGAACACCTTGGTCGACACCTCCGGTTACACACGGGAGG 163
Db 80 TCCTGAACCTGGCCTACGAATCCAATCACTGATGACCTGTCTCGCTACGTTCCAAA 139
Qy 164 TCTCCGAGGAGGTGAGCTCCAGCTGAACCCAAATCTTCCATTCGACTTCAAGCTGGGT 223
Db 140 TCAACATCGGTTCTAAAGTTAACTTCGATCCGATCGACAAGATCAGATCCAGCTG --- 195
Qy 224 CTCCGGTGAGGACAGGTPAAGGTGATCGTACCCAGAACGAGAACATCGTCTTCAACT 283
Db 196 --TTCAATCTGGAATCTTCCAAAATCGAAGTTATCTCTGAAAGATGCTATCGTATACA 253
Qy 284 CCATGTACGAGTCTCTTCCATCTCTTCTGGATCAGATCAACAACTGGGTCTTCCAAC 343
Db 254 CTAATGACGAAAATCTTCCACCTCTTCTGGATCGTATCCGAAATTAATCAACTCCA 313
Qy 344 TGCC-----AGTTACACCATCATCGACTCCGTCGATCAAGAACAACTCCGTTTGGTCCA 394
Db 314 TCTCTGAAACAATGAATACACCATCATCAACTGCATGCAATGGAACAAATCTGTTGGTGA 373
Qy 395 TCGGTATCATCTCCAACCTTCTTCCCTGCTTCACTGAGGACGAGGAGGACTCCGAGCAGT 454
Db 374 TATCTGTAATACGGTGAATCATCTGGACTCTGGAGGACACTCAGGAATTAACACAGC 433
Qy 455 CCATCAACTTCTCTAGACATCTCCAACAACGCTCTGTTTACACAA---GTGTTTCT 511
Db 434 GTGTTGTAATCAAACTCTCAGATGATCAACATCTCTGACTACATCAATCGTGGATCT 493
Qy 512 TCGTCACTGTCACCAACATCATGTTGGTAAACATGAAGATCTACATCAACGGTAAAGCTGA 571
Db 494 TCGTTACCATCACCACAAATCGTCTGAATAAATCCAAAATCTACATCAACGGCCGCTGA 553
Qy 572 TCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCCAGACCATCACTCTCG 631
Db 554 TCGACCAAGAACCGATCTCCAATCTGGGTAACATCCACGCTTCTAATAACATCATGTTC 613
Qy 632 AGATCAACAAGATCCACAGACACCGGTTCTGATCACCCTCCGACTCCGACACATCAACA 691
Db 614 AACTGGACGGTTGCTGGACACTC-----ACCGTACATCT 649
Qy 692 GGATCCGTGACTTCTACATCTTCGCAAGGAGTTGGACGTAAGGACATCAACATCTCTGT 751
Db 650 GGATCAATACTTCAATCTGTTCCACAAAGACTGAACGAAAAAGAAATCAAGACCTGT 709
Qy 752 TCAACTCCTTGCAGTACACCAACGTCGTCAGGACTACTCGGGTACGACCTTGAGATACA 811
Db 710 ACGAACCCAGTCCCAATTTCTGTTATCTCTGAAAGACTCTTGGGTGACTTACCTGCA 769
Qy 812 ACAAGGAGTACTCATGTTCAACATCGACTTCTGAAACAGATACATGTACGCCAAC 867
Db 770 ACAACCGTACTACTGCTGAAATCTGTACGATCCGACAAATACGTTGAGTCAAC 825

RESULT 13
AAAS4483
ID AAAS4483 standard; DNA; 1323 BP.
XX AAAS4483;
AC AAAS4483;
XX 11-APR-2001 (first entry)
DE Botulinum toxin heavy chain C-terminal coding sequence (serotype A).
KW Botulinum toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;

```



```

PR 14-APR-1995; 95US-0422711.
XX (OPHI-) OPHIDIAN PHARM INC.
XX Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
XX Williams JA;
XX WPI: 1996-230603/23.
XX P-PSDB: AAR95009.
XX Fusion proteins comprising non-toxin protein and part of toxin
XX useful to form anti-toxins against Clostridium botulinum type A, and
XX C. difficile type toxins, and to treat C. difficile intoxication,
XX partic. diarrhoea
XX Example 24; Page 340-342; 434pp; English.
XX A nucleotide sequence (AAR29246) present in vector pETHisa encodes
XX the phisBot fusion protein (AAR95009) comprising a polyhistidine
XX affinity tag and fragment C (see also AAR95008) of the Clostridium
XX botulinum type A neurotoxin. The phisBot protein was expressed
XX in Escherichia coli as a soluble protein and was purified by
XX metal chelate affinity chromatography to obtain a product free
XX of endotoxin contamination that may be useful as an immunogen
XX in vaccine compens.
XX Sequence 1402 BP; 420 A; 360 C; 260 G; 362 T; 0 other;
XX
XX Query Match 12.4%; Score 169.4; DB 17; Length 1402;
XX Best Local Similarity 54.4%; Pred. No. 2.2e-27;
XX Matches 456; Conservative 0; Mismatches 341; Indels 42; Gaps 4;
XX
XX 41 CCAACAACCTCCTGTTGAAGGACATCATCAACGAGTACTTCAACACATCAACGACTCCA 100
XX 68 CTAGCATGGCTGCTGCTGCTACCTTCACTGAATACATCAAGAACATCATCAATACCT 127
XX 101 AGATCTCTGCTCCGACAAACGGTAAGNACACCTTGTGACACACCTCCGGTTACAGCGG 160
XX 128 CCATCTCTGAAGCTCGGTGAGATCCAAATCAACCTGATCGACCTGTCTCGCTACGGTTCA 187
XX 161 AGGTCCTCGAGGAGGTGAGTCCAGCTGAACCCCAATCTCCCATTCGACTTCAAGCTGG 220
XX 188 AAATCAACATCGGTTCTAAAGTTAACTTCGATCCGATCGACAGATGATGATCCAGCTG- 246
XX 221 GTTCCTCGGTGAGGACAGAGTAAAGTCAATCGTACCCAGAACAGACATCGCTTACA 280
XX 247 -----TTCAATCTGGAATCTTCCAAAATGAAGTTATCTCTGAGAAATGCTATCGTATACA 301
XX 281 ACTCATGTACGAGTCTCTTCCATCTCTTCTGGATCAGAATCAACAAGTGGTCTCCA 340
XX 302 ACTCTATGTAGAAACTTCTCCACCTCTCTTGGATCCGATCCGGAATACTTCACT 361
XX 341 ACTTGCC-----AGGTTACACATCATCATCGATCCGTTCAAGAACACTCCGGTTGTT 391
XX 362 CCATCTCTGAAACAATGAATCAACCATCATCACTCACTGATGGAAGAAACAATCTGTTGGA 421
XX 392 CCATCGGTATCATCTCCAACCTCTCTGCTCTTCACTGAGCAGAGAGAGCTCGAGC 451
XX 422 AAGTATCTCTGAATAGGTGAATATCTCTGAGTCTGAGGACATCTGAGGAATCAAC 481
XX 452 AGTCCATCAACTTCTCTGACATCTTCCAAACAGCTCTCTGTTTACAACA---GTGTT 508
XX 482 AGCGTGTGTATTCAAAATCTCTCAGATGATCAACATCTCTGACTACATCAATCGGTGA 541
XX 509 TCTTCGTCACGCTCACCACAAACATGATGGTGAATGATGAGATCTACATCAACGGTAAAC 568
XX 542 TCTTCGTTACCATCACCACAAATGCTGTAATCACTCCAAATCTACATCAACGGCGGTC 601
XX 569 TGATCGACACCATCAAGGTCAAGAGTTGACCGGTATCACTTCTCCAAGACATCACT 628
XX 602 TGATCGACCAAGAACCGATCTCCAATCTGGGTAACTCCACGCTTCTAATACATCATGT 661
XX 629 TCGAGATCAACAAGATCCGACAGACCGGTTGATCACTCCGACTCCGACACATCAACA 688
XX
XX 662 TCAAACCTGGACGGTTGCTGTGACATC-----ACCCCTACA 697
XX 689 TGTGGATCCGTGACTTCTATCATCTTCCCAAGGAGTTGACGCGTAAGGACATCAACATCC 748
XX 698 TCTGGATCAATATCTTCAATCTCTGACAAAGAACTGAACGAAAAAGAAATCAAGACC 757
XX 749 TGTCAACTCTTTCAGTACACCAACGCTGTCAGGACTACTGGGTACGACCTGAGAT 808
XX 758 TGTACGACAAACACAGTCCAAATTCCTGATGAGACTTCTGGGTGACTACCTGCGAGT 817
XX 809 ACACAAAGGAGTACTACTATGTCATGTCATGATGACTTCTTGAACAGATACATGTACGCCAAC 867
XX 818 ACGCAAAACCGTACTACTATGCTGATCTGTAGATCCGAAACAATACGTTGACGTCAAC 876
XX
XX RESULT 15
XX AAV30572
XX ID AAV30572 standard; DNA; 1402 BP.
XX AC AAV30572;
XX XX
XX DT 07-DEC-1998 (first entry)
XX XX
XX DE Clostridium botulinum toxin A fragment C gene in phisBot.
XX XX
XX KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; Immunogen;
XX botulism; ds.
XX XX
XX OS Clostridium botulinum serotype A.
XX OS Synthetic.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 1..1389
XX FT /*tag= a
XX XX
XX DN WO9808540-A1.
XX XX
XX PD 05-MAR-1998.
XX XX
XX PF 28-AUG-1997; 97WO-US15394.
XX XX
XX PR 28-AUG-1996; 96US-0704159.
XX XX
XX PA (OPHI-) OPHIDIAN PHARM INC.
XX XX
XX PI Thalley BS, Williams JA;
XX XX
XX PS WPI: 1998-230234/20.
XX P-PSDB: AAW68390.
XX XX
XX PT Host cell containing recombinant expression vector encoding
XX Clostridium botulinum type B or E toxin - useful to treat humans
XX and other animals at risk of intoxication with clostridial toxin
XX Example 24; Page 265-267; 428pp; English.
XX XX
XX CC This is the DNA sequence of the Clostridium botulinum serotype A
XX toxin C fragment gene contained in plasmid phisBot. The encoded
XX toxin A polypeptide (see AAW68390) has a histidine-tagged
XX N-terminal extension. The vector was used to express native
XX (i.e. non-fusion) soluble C fragment in Escherichia coli host
XX cells. The invention relates to recombinant proteins derived from
XX C. botulinum toxins. Methods are provided which allow for the
XX isolation of soluble recombinant proteins free of significant
XX endotoxin contamination. Preferred hosts for production of
XX recombinant proteins are E. coli, insect cells and yeast cells.
XX The recombinant toxins are used as immunogens for the production
XX of vaccines and antitoxins that are useful in the treatment of
XX humans and animals at risk of intoxication with clostridial toxin.
XX Sequence 1402 BP; 420 A; 360 C; 260 G; 362 T; 0 other;
XX

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:20:20 ; Search time 57 Seconds  
(without alignments)  
7376.383 Million cell updates/sec

Title: US-09-910-186A-9  
Perfect score: 1371  
Sequence: 1 gaattacagatgaccatcccc.....tctccagtagtaagaattc 1371

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	169.4	12.4	1402	1	US-08-480-604A-25
2	169.4	12.4	1402	2	US-08-405-496A-25
3	169.4	12.4	1402	4	US-08-915-136-25
4	168.8	12.3	1330	1	US-08-480-604A-22
5	168.8	12.3	1330	2	US-08-405-496A-22
6	168.8	12.3	1330	4	US-08-915-136-22
7	136	9.9	5163	3	US-08-700-651-1
8	136	9.9	5163	3	US-08-928-361B-4
9	136	9.9	5318	3	US-08-700-651-2
10	136	9.9	5318	3	US-08-928-361B-3
11	115.4	8.4	5511	3	US-08-928-361B-2
12	115.4	8.4	7334	3	US-08-928-361B-1
13	95	6.9	4483	4	US-08-961-527-363
14	93.4	6.8	2004	1	US-08-471-033-18
15	93.4	6.8	2004	2	US-08-471-044-18
16	93.4	6.8	2004	2	US-08-463-483A-18
17	93.4	6.8	2004	2	US-08-471-046A-18
18	93.4	6.8	2004	2	US-08-470-566B-18
19	93.4	6.8	2004	2	US-08-469-334-18
20	93.4	6.8	2004	3	US-09-300-529-18
21	93.4	6.8	2576	1	US-08-471-033-35
22	93.4	6.8	2576	2	US-08-471-044-35
23	93.4	6.8	2576	2	US-08-463-483A-35
24	93.4	6.8	2576	2	US-08-471-046A-35
25	93.4	6.8	2576	2	US-08-470-566B-35
26	93.4	6.8	2576	2	US-08-469-334-35
27	93.4	6.8	2576	3	US-09-300-529-35

28	93.4	6.8	2655	1	US-08-471-033-17
29	93.4	6.8	2655	1	US-08-471-033-26
30	93.4	6.8	2655	2	US-08-471-044-17
31	93.4	6.8	2655	2	US-08-471-044-26
32	93.4	6.8	2655	2	US-08-463-483A-17
33	93.4	6.8	2655	2	US-08-463-483A-26
34	93.4	6.8	2655	2	US-08-471-046A-17
35	93.4	6.8	2655	2	US-08-471-046A-26
36	93.4	6.8	2655	2	US-08-470-566B-17
37	93.4	6.8	2655	2	US-08-470-566B-26
38	93.4	6.8	2655	2	US-08-469-334-17
39	93.4	6.8	2655	2	US-08-469-334-26
40	93.4	6.8	2655	3	US-09-300-529-17
41	93.4	6.8	2655	3	US-09-300-529-26
42	93.4	6.8	4031	1	US-08-471-033-49
43	93.4	6.8	4031	2	US-08-471-044-49
44	93.4	6.8	4031	2	US-08-463-483A-49
45	93.4	6.8	4031	2	US-08-471-046A-49

ALIGNMENTS

RESULT 1  
US-08-480-604A-25  
Sequence 25, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,604A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027

DBD 818 ACGACAAACCTACTAGATCTGTACGATCCGACAAATACGTTGACGTCAAC 876

RESULT 2

US-08-405-496A-25

; Sequence 25, Application US/08405496A

; Patent No. 5919665

; GENERAL INFORMATION:

; APPLICANT: WILLIAMS, JAMES A.

; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

; TITLE OF INVENTION: NEUROTOXIN

; NUMBER OF SEQUENCES: 30

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA

```

ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1386
US-08-405-496A-25

Query Match      12.4%; Score 169.4; DB 2; Length 1402;
Best Local Similarity 54.4%; Pred. No. 8e-29;
Matches 456; Conservative 0; Mismatches 341; Indels 42; Gaps

QY    41  CCACACTCCCTGTTGAGGACATCATCAAGCAGTACTTCACAACATCAAGACTCCA 100
Db     68  CTAGCATGGCTCGCTGCTGTCTACCTTCAGTAATACATCAAGAATCATCATCAATACCT 127
QY    101 AGATCCCTGCCTCCGACAGCCGTAGAACACCTTGCTGGTAGCACCTCCGGTTTACAACGCCG 160
Db     128 CCATCCTCGAACCTGCGGTAGCAATCCAACTCACTGATCGACTCTCTCGGTAGCTTCCA 187
QY    161 AGGTCCTCGAGAGGGTGACGTCCAGCTGAACCCAAATCTCCCATTGCACTTCAAAGCTGG 220

```

Query Match	12.4%;	Score 169.4;	DB 2;	Length 1402;
Best Local Similarity	54.4%;	Pred. No. 8e-29;		

41	QY	CCACAACTCCCTGTTGAGGAGCATCTCAGGAGTACTTCAACACATCAAGCATCA	100
68	Db	CTAGCATGGTGCCTGCTGTACCTTCACTGAATACATCAAGCATCTCATTA	127
101	QY	AGATCTCTCCCTCGAGAACCGTAAAGAACCTTGGTCGACCTCCGGTTTCAACGCG	160
128	Db	GCATCTCTGAACCTGGGTAGAAATCAATCACTGATCGACTCTCTCGCTACGCTTCA	187
161	QY	AGGCTCCGAGAGGGGTGAGTCCAGCTGAACCCAAATCTCCCAATTCGACTTCAAGCTGG	220

Db 188 AAATCAACATCGGTCTTAAAGTTAACTCGATCGACGAAGAATCAGATCCAGCTG- 246  
Qy 221 GTTCTCGGTGAGCAGAGAGGTAAAGTTCATCGTCCACCAAGAGAGATCGTCTACA 280  
Db 247 -----TTCAATCTGAACTCTCCAAATCGAAGTTATCTCGAAGATGCTATCGTATACA 301  
Qy 281 ACTCGATGATAGATCTTCTCCATCTCTTGGATCAGATCAACAGAGTGGGTCTCA 340  
Db 302 ACTATGATAGAAATCTTCCACCTCTCTTGGATCGGTATCCGGAATATCTCAACT 361  
Qy 341 ACTTGCC-----AGGTTACACCATCATCGACTCCGTCAGAACCACTCCGGTGGT 391  
Db 362 CCACTCTCTGAACAATGATACACCATCATCACTGCAATGGAACAATCTGTTGA 421  
Qy 392 CCACTGGTATCATCTCAACTCTCTGCTTTCACCTGGAAGAGAGAGGACTCCGAGC 451  
Db 422 AAGTATCTCTGAATAGGTTAAATCATCTGACTCTGCAGGACACTCAGGAATCAAC 481  
Qy 452 AGTCCATCAACTTCTCTAGACATCTCCAAACAGGCTCTGTTTACAA--GTGGT 508  
Db 482 AGCGTGTGTTATCAATACTCTCAGATGATCAACATCTCTGACTACATCAATCGTGA 541  
Qy 509 TCTTGTCACCGTCCACCAACATGATGGGTACATGATGATCTACATCAACGGTAAAC 568  
Db 542 TCTTGTTTACCATCACCACAAATGCTGTAATCACTCAAAATCTACATCAACGGCGTC 601  
Qy 569 TGATCGACACCATCAAGTCAAGGTTGACCGGTATCACTTCTCAAGACCATCACT 628  
Db 602 TGATCGACCAACCATCTCAATCTGGTACATCCAGCTTCTATATACATCATGT 561  
Qy 629 TCGATCAACAAGATCCAGACACACCGGTCTGATCACTCCGACTCCGACACATCAACA 588  
Db 662 TCAAACTGGAGGTTGCTGACACTC-----ACCGTACA 697  
Qy 689 TGTGATCGGTGACTTCTACATCTTCCCAAGAGTTGGAGGTGAGGACATCAACATCC 748  
Db 758 TGTAGCAACACAGTCCATCTGTTATCTGTTATCTGAAAGACTTCTGGGTGACTACCTSCAT 817  
Qy 809 ACAACAGGATCTATGATGTCACATCACTGATGATGATGATGATGATGATGATGATGAT 867  
Db 818 ACGAACAACGCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876

RESULT 3

us-09-915-136-25  
Sequence 25, Application US/08915136  
Patent No. 6290960

GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,136  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,604  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1402 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1386  
US-08-915-136-25

Query Match 12.4%; Score 169.4; DB 4; Length 1402;

Best Local Similarity 54.4%; Pred. No. 8e-29; Mismatches 341; Indels 42; Gaps 4;  
Matches 456; Conservative 0;

Qy 41 CCACAACTCCCTGTTGAGGACATCATCAACGAGTACTTCAACACATCAACAGCTCCA 100  
Db 68 CTAGCATGGCTGCTGCTGTCTACCTTCACTGAATACATCAAGAACATCAATACCT 127  
Qy 101 AGATCCTGTCCTGCAGAACCGTAAGAACACCTTGTGTCACACCTCCGGTTACAAGCGG 160  
Db 128 CCATCCTGAACCTGCGTACGAATCCATCACTGATGACCTGTCTCGTACGTTCCA 187  
Qy 161 AGGTCCCGAGGAGGTGACGTCCAGTGAACCCAAATCTTCCATTTCGACTTCAAGCTGG 220  
Db 188 AATCAACATCGGTTCTAAAGTTAACTTCGATCCGATCGACAGAATCAGATCCAGCTG- 246  
Qy 221 GTTCTCCGGTGAGGACAGAGTAAAGTATCTGTCACCCAGACAGAGACATCGTCTACA 280  
Db 247 -----TTCAATCTGGAATCTTCCAAAATCGAAGTTATCTGGAATGCTATCGTATACA 301  
Qy 281 ACTCATGTACGAGTCTCTTCATCTCTCTGATCAGATCAAGATCAACAAGTGGTCTCCA 340  
Db 302 ACTATGTACGAATCTTCTCCACTCTCTTGGATCGGTATCCGGAATATCTTCAACT 361  
Qy 341 ACTTGCC-----AGGTTACACCATCATCGATGCTGCTCAAGAACAACTCCGGTGGT 391  
Db 362 CCATCTCTGTAACAATGAAATACCATCATCAACTGATGGAAGAACAACTTCTGGTGA 421  
Qy 392 CCATCGGTATCATCTCAACTTCTGCTTTCACCTGGAAGAGAGAGGACTCCGAGC 451  
Db 422 AAGTATCTGAACTACGGTGAATCATCTGGACTCTGCAGGACACTCAGGAATCAAC 481





QY 772 AACGTCGTCAGGACTACTGGGTAAACGACCTCGAGATACAAAGAGGAGTACTACATGGTC 831  
DB 709 GGTATCTGAAAGACTTCTGGGTGACTACTCGCTAGTACGAAACCGTACTACATGCTG 768  
QY 832 AACATCGACTACTTGAACAGATACATCATGCGCAAC 867  
DB 769 AATCTGTACGATCGGACAAATAGTTGACGTCAAC 804

RESULT 5  
US-08-405-496A-22  
; Sequence 22, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/405,496A  
; FILING DATE: 16-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-01308  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1330 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1314  
US-08-405-496A-22

Query Match 12.3%; Score 168.8; DB 2; Length 1330;  
Best Local Similarity 54.8%; Pred. No. 1.1e-28;  
Matches 447; Conservative 0; Mismatches 327; Indels 42; Gaps 4;  
QY 64 ATCATCAAGGACTTCAACAACTCAAGACTCCAGATCTGTCCTGCGAGAACCGT 123  
DB 19 ACCCTCACTGATACATCAGACATCATCACTCCATCTGACCTGCGCTACGAA 78

QY 124 AAGAACACCTTGGTGGACACCTCGGTTTACAAACCGGAGGTCTCCGAGAGGGTGACGTC 183  
DB 79 TCCAATCACCTGTAGTACGACCTGTCTCGTACGCTTCAAAAATCAACATCGGTTCTAAAGTT 138  
QY 184 CAGTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTTCCCTCCGGTGAGGACAGAGT 243  
DB 139 AACTTCGATCGATCGACAAAGATCAGATCCAGCTG-----TTCAATCTGGAATCTTCC 192  
QY 244 AAGTCATCTGTCACCCAGAGAGAGACATCGTCTACAATCTCCATGTCAGAGTCTTCTCC 303  
DB 193 AAAATCGAAGTTATCTCTGGAAGATGTATCGTATACAACATCTATGTACGAAACATCTCTCC 252  
QY 304 ATCTCTTCTGGATCAGAATCAACAGTGGTCTCAACATCTG-----AGGTTAC 354  
DB 253 ACCCTCTCTGGATCCGATATCCCGAATATCTCAATCTCTCTGACATGATATC 312  
QY 355 ACCATCATCGACTCCGTCAGAGAACTCCGTTGGTCCATCGGTATCATCTCCCAATTC 414  
DB 313 ACCATCATCAACTGCATGGAAGAACTCTGTTGGAAGTATCTCTGAATACGTTGAA 372  
QY 415 CTGCTCTCACTCCGTCAGAGAGAGGACTCCGAGCAGTCCATCAACTCTCTCTCTACGAC 474  
DB 373 ATCATCTGGACTCTGAGGACTCAGGAATCAACAGCGTCTGTATTCAAATCTCT 432  
QY 475 ATCTCCAAACACGCTCTGTTTACAAACAA---GTGGTTCTTCTGTCACCGTCAACCAAC 531  
DB 433 CAGATGATCAACATCTCTGACTATCAATCGCTGGATCTTCTGTTACCATCAACAACT 492  
QY 532 ATGATGGTAACTGAGATCTATCAACAGGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAG 591  
DB 493 CGTCTGAATAACTCCAAAATCTATCAACAGGCGCTCTGATCGACGAGAAACCGATCTCC 552  
QY 592 GAGTTGACCGGTATCAACTCTCCAGAGACCATCACCTTCGAGATCAACAGATCCAGAC 651  
DB 553 AATCTGGTAACTCCAGCTTCTATAACATCATGTTCAAACTGGAGCGTTGTCGTGAC 612  
QY 552 ACCGGTCTGATCACTCCGACTCCGACATCAACATCAAGTGGATCGGTGACTCTTCAATC 711  
DB 613 ACTC-----ACCGGTACTCTGGATCAATCAATCAATCAATCAATCAATCAAT 648  
QY 712 TTGCGCAAGGAGTTGGACGTTAAGGACATCAACATCTGTTCAACTCTCTGAGTACAC 771  
DB 649 TTGCAAAAGACGTCAGAGAAAGAAATCAAGACCTGATCAACACCGATCTCTTCT 708  
QY 772 AACGTCGTCAGGACTACTGGGTAAACGACCTCGAGATCAACAGAGGAGTACTACATGGTC 831  
DB 709 GGTATCTGAAAGACTTCTGGGTGACTACTGCTAGTACGAAACCGTACTACATGCTG 768  
QY 832 AACATCGACTACTTGAACAGATACATCATGTCGCAAC 867  
DB 769 AATCTGTACGATCGGACAAATAGTTGACGTCAAC 804

RESULT 6  
US-08-915-136-22  
; Sequence 22, Application US/08915136  
; Patent No. 6290960  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTI-TOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA

313	ACCATCATCAACTGCGATGGAAAAAACAATTCTGGTTGGAAAGTAATCTCTGCACTACGCTGA	378
415	CTGGTCTTCACCCCTGAAGCAGAACAGGAGGACTCCGAGCAGTGCCATCAACTTCTCCTTAGCAC	474
373	ATCATCTGGACTCTGCAGGACACTCAGGAAGAATAACAAACGGGTGTGTATCAAAATACTCT	432
475	ATCTCCAACAACGCTCCTGGTTACAACAA---GTGGTTCTTCGTACCGGTACCAACAAC	531
433	CAGATGATCAATCATCTCTGACTACATCAATCGCTGGATCTTCGTTAGCATCACCACAAT	492
532	ATGATGGGTAAACATGAAGATCTACATCAACGGTAAGCTGATCGACACCATCAAGGTCAAG	591
493	CGTCTGAATAACTCCAAAATCTACATCAACGCCGCTCTGATCGACCAAGAACCGATCTCC	552
592	GAGTTGGCCGGTATCAACTTCTCCAGACCATCACCTTCGAGATCACAGAATCCCAGAC	651
553	AATCTGGGTAAACATCACCGCTCTTAATACATCATGTTTCAAACCTGGACGGTGTGCTGAC	612
652	ACCGGCTGTGATCACTCCGACTCCGACACATCAACATGTGGATCGGTGACTTCTATCATC	711
613	ACTC-----ACCGTACATCTGGATCGAAATACTTCAATCTG	648
712	TTCGCCAAGGAGTTGGACGGTAAGGACATCAACATCTCTGTTCAACTCCTTGCAGTACACC	771
649	TTCGACAAAGAACTGAACGAAAAGAAATCAAAGACCTGTACGCAACACCAAGTCCAATCTC	708
772	AACGTCGTCAAGGACTACTGGGGTAACGACCTGAGATACACACAGGAGTACTACATGGTC	831
709	GGTATCTTGAAGACHTCTGGGGTGACTCTTGATGTACGCAACACCGCTACTACATCTGCTG	768
832	AACATGACACTTGTAAACAGATACATGTACGCCAAC	867
769	AATCTGTACGATCCGAACAAATACGTTGACCTCAAC	804

## RESULT 7

US-08-700-651-1  
Sequence 1, Application US/08700651B

APPLICANT: RHEONANT, KILBURN, G. C.  
 TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
 TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum

```

; TITLE OF INVENTION: INFECTIONS
;
; FILE REFERENCE: 480.19-4 (HV)
;
; CURRENT APPLICATION NUMBER: US/08/700.651B
;
; CURRENT FILING DATE: 1997-08-14
;
; EARLIER APPLICATION NUMBER: 08/415.751
;
; EARLIER FILING DATE: 1995-04-03
;
; NUMBER OF SEQ ID NOS: 15
;
; SOFTWARE: PatentIn ver. 2.0
;
; SEQ ID NO 1
;
; LENGTH: 5163
;
; TYPE: DNA
;
; ORGANISM: Cryptosporidium parvum
;
; US-08-700-651-1

```

	Query Match	9.9%;	Score 136;	DB 3;	Length 5163;
	Best Local Similarity	50.6%;	Pred. No. 2.9e-21;		
	Matches 328;	Conservative 0;	Mismatches 320;	Indels 0;	Gaps
541	AAATGGAAGATCTACATCAACGGTGAAGCTGATCGACACCATCAAGGTCAGAGGATTCACC	600			
577	AACAACAACAACAACACTACTACTACTACTACTAGCAACAACAACAACAACAGCAAC	636			
601	GGTATCAACTTCCAGACCATCACTTCGAGATCAACAAGATCCCAAGACACCGGTCTG	660			
637	AACAACAACACTACTACTACTACTACTACTAGCAACAACAACAACAACAACAAC	696			
661	ATCACTTCGGACTCGGACACATCAACATGTGGATCGGTGACTTCTACATCTTCGCCAAG	720			







MOLECULE TYPE: DNA (genomic)  
US-08-928-361B-1

Query Match  
Best Local Similarity 8.4%; Score 115.4; DB 3; Length 7334;  
Prior Application Data: 49.0%; Pred. No. 1.2e-16;  
Matches 308; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY 606 CAATCTCCAGACATCACTCGGATCAACAGATCCAGACACCGCTCTGATCAC 665  
DB 2621 CTACTACTAGCAACCAACCAACAGACCAACCAACCAACCAACCAACCAACCA 2680

QY 666 CTCGACTCCGACATCAACATGATGATCGGTGACTTCTACATCTTCGCGCAAGGAGTT 725  
DB 2681 CTACTACTAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 2740

QY 726 GGACGCTAGGACATCAACATCTGTTCAACTCTTGGAGTACACCAACCAACCAACCA 785  
DB 2741 CTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 2800

QY 786 CTACTGGGTAGGACCTGAGATCAACAAAGGAGTACTACATGTTCAACATCGACTACTT 845  
DB 2801 CAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 2860

QY 846 GAACAGATACATGATGACCACTCAGACAGATCGTCTTCAACACCAACCAACCAACCA 905  
DB 2861 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 2920

QY 906 CGACTTCAACGAGGTTTACAAAGTATCATCAAGCGTATCAGAGGTATCAACCAACCAAC 965  
DB 2921 CCACACTCCAGAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 2980

QY 966 CAGAGTCAGAGGTGGTACATCTGTTACTTTCGACATGATCTATCAACCAACCAACCA 1025  
DB 2981 CCACGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 3040

QY 1026 CCGTTCATGAAGACGAGACCATGTACCGGACACCAACCAACCAACCAACCAACCA 1085  
DB 3041 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3100

QY 1086 CATCGGTCTGCGTGACGACCAACGACATCAACGACCAACCAACCAACCAACCAACCA 1145  
DB 3101 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3160

QY 1146 AATGAACACACTTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAGAA 1205  
DB 3161 CAACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 3220

QY 1206 CATCTCCGTTATGTTTCCATCGGTACCT 1234  
DB 3221 CCACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 3249

RESULT 13  
US-08-961-527-363  
Sequence 363, Application US/08961527  
Patent No. 6420135

GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961.527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 363:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4493 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-363

Query Match  
Best Local Similarity 6.9%; Score 95; DB 4; Length 4483;  
Matches 558; Conservative 0; Mismatches 665; Indels 24; Gaps 4;

QY 29 TCTTCTCTTACACCAACCACTCCCTGTTGAAGGACATCATCAACGAGTACTTCAACAA 88  
DB 2362 TCAGGCTCTGAATCGGCTCAACGAGTGCCTCAGCCTCAGCAAGTACTAGTGCATCAGCA 2421

QY 89 TCACGACTCCAGATCTCTGCTCCCTGAGACCGGTAGAACACCTTGGTCGACACTCCG 148  
DB 2422 TCAGCATCAACGAGTGCCTGCTTCAGCAAGTACCGAGCCTCAGCTTCAGCAAGCACC 2481

QY 149 GTTACAAACCCGAGGTCTCCGAGGAGGTGAGCTCCAGCTGAACCCCAATCTTCCCATTCG 208  
DB 2482 AGTGGCTCAGCTCAGCAAGTACCAAGCGCTCAGCCTCAGCAAGCACCAGTGCCTCAGCT 2541

QY 209 ACTTCAAGCTGGTTCCTCCGGTGAAGGACAGAGTAAAGTATCTGTCACCCAGAACAGA 268  
DB 2542 TCAGCAAGTACCAAGTGCCTGCTTCAGCCTCAGCCTCAGCAAGTGCCTGCTTCAGCAAGTACC 2601

QY 269 ACATCTCTTACAACTCCATGATGACGAGTCTCTCCATCTCCTTCTGATCAGAAATCAACA 328  
DB 2602 TCAGCTCTGATGACGATCAACGAGTGCATCAGCTTCAGCATCAACAAAGTGCCTCAGCT 2661

QY 329 AGTGGGTCTCAACTTCCAGGTTACACCATCATCGATCCGTCAGCAAGCAACTCCGGTT 388  
DB 2662 TCAGCAAGTACCAAGTGCCTGCTTCAGCCTCAGCCTCAGCAAGTGCCTGCTTCAGCTCAACC 2721

QY 389 GGTCCATCGGTAT---CATCTCCAACTTCCCTGGTCTTCAACCTGAAGCAGCAAGGAGCT 445  
DB 2722 AGTGGCTCTGATCAGCATCAACAGTGCCTCGGCTTCAGCAAGCACCAGTGCCTCGGCT 2781

QY 446 CCGAGCAGTCCATCAACTTCTCTAGCAGATCTCCCAACAGCTCCCTGGTTTACAAAGT 505  
DB 2782 TCAGCAAGTACTAGTGCATCGCTTCAGCCTCAGCAAGTGCCTGCTTCAGTTCGGCAGCA 2841

QY 506 GGTCTCTGCTCAGCTCAGCAACAA-----CATGATGGTAAACATGAAGATCTACA 556  
DB 2842 AGTGGCTTCGCTCAGCATCAACAGTGCCTCAGCCTCAGCAAGCACCAGTGCCTGCTGAA 2901

QY 557 TCAACGGTAAAGTGTATCGACACCAATCAAGGTCAAGGATGACCGGTATCAACTTCTCCA 616  
DB 2902 TCTGCATCAACCAAGTGCCTGCTTCAGCCTCAGCAAGTGCCTGCTTCAGCTTCGACG 2961

QY 617 AGACCATCACTTCCGAGATCAACAGATCCCGACACCGGCTCTGATCATCCTCCGACTCCG 676  
DB 2962 AGTGGCTTCGCTCAGCATCAACAGTGCCTCAGCCTCAGCAAGCACCAGTGCCTGCTGAA 3015

QY 677 ACAACATCAACATGATGATCGGTGATCTTCTACATCTTCGCAAGGAGTTCGACGGTGAAG 736  
DB 3016 TCAGCGTTCAGCTTCGCGCTCAACCAAGTGCCTGCTTCAGCAAGCACCAGTGCCTGCTGAA 3075

QY 737 ACATCAACATCCTGTTCAACTCCTTGCAGTACACCAACAGTCTGTCAGAGGACTACTGGGTA 796  
Db 3076 TCAGCAAGATATCTCAGGCTGTGAATCGGCATCAACAGAGTGCCTCGGCCTCAGCAAGCGCA 3135  
QY 797 ACGACCTGAGATACAAAGAGTACTGATGTTCAACATCGACTACTTGAACAGATACA 856  
Db 3136 AGTACCTCAGGCTCAGC-----TTCCGCTCAACAGTGCCTCGGCTTCAGCAAGCACA 3189  
QY 857 TGTAGGCCAATCCAGACAGATGCTTTCAACACCCAGACGTAACAAACAGAGTTCAAGC 916  
Db 3190 AGTGGCTCAGCCTCAGCAAGTATCTCAGCGTCTCAATCGGCATCAACAGAGTGGCTGAG 3249  
QY 917 AGGTTAAGATCATCATCAAGAGTATCAGAGTACACCAACAGCAGCAGAGTACAGAG 976  
Db 3250 TCAGATCAAGAGTACTGATGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3309  
QY 977 GTGTGACATCCTGATCTGACATGACTATCAACAAAGGCTTCAACCTTTTCATGA 1036  
Db 3310 AGTGGCTCAGCCTCAGGATCGACAAAGCGCTCAGCTTCAGCAAGTACCAGTGTCTCAGCC 3369  
QY 1037 AGAAGCAGACCATGTAGCCGACAAACACATCCACCGAGGACATCTAGCCATCGGTCTGC 1096  
Db 3370 TCAGGCTCGCAAGTGGTGGCTCAACAGTGCATCTGAATCGGCATCAACAGCAGTGG 3429  
QY 1097 GTGACGACCAAGAGCATCAACAGACATCATCTTCCAGATCCACCAATGAACACA 1156  
Db 3430 TCAGCCTCAGCAAGTACTAGTGCATCAGCTTCAGCATCAACAGAGTGCATCGGCTTCAGCA 3489  
QY 1157 CTTACTACTAGCTTCCAGATCTTCAAGTCCACTTCAAGCTGAGAGATCTCCGGTA 1216  
Db 3490 TCAACCAAGTGGCTCGGCTTCAGCGTCAACAGTGGCTCAGCTTCAGCAAGTACCAAGTGT 3549  
QY 1217 TGTGTTCACTCGGTACCTACAGATTCGCTGGGTGGTGGTACTGGTAC 1263  
Db 3550 TCAGTCTCAGCATCAACAAGTGTTCAGCCTCAGCATCAGCAAGTGC 3596

## RESULT 14

US-08-471-033-18

Sequence 18, Application US/08471033

Patent No. 5770696

## GENERAL INFORMATION:

APPLICANT: Warren, Gregory W  
APPLICANT: Kozel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentlin Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,033  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: P-40,403  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2004 base pairs  
TYPE: nucleic acid  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..2004  
OTHER INFORMATION: /note= "Maize optimized DNA"  
OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"  
US-08-471-033-18

Query Match 6.8%; Score 93.4; DB 1; Length 2004;

Best Local Similarity 44.3%; Pred. No. 6,1e-12;

Matches 529; Conservative 0; Mismatches 651; Indels 15; Gaps 3;

QY 51 CCTGTTGAAGGACATCATCAACGAGTACTTCAACAACATCAACGACTCCAAAGATCTCTGTC 110  
Db 378 CGTGAGGCGCGCATCGTCCCAAGGCGATCAGCTTCGGGCTGAGCGTGAATACCGCA 437  
QY 111 CCTGCAGAACCGTAAAGAACACCTTTGGTCGACACCTCCGGTTACACGCCGAGGTCTCCGA 170  
Db 438 CAGCGAGACCGTGGCCCGAGGTGGGCGACGACGCGGCAACACGAGCGCATCTCAACAC 497  
QY 171 GGAGGTTGAGTCCAGCTGAACCCCAATCTCCCATTCGACTTCAAG--CTGGGTTCCTC 227  
Db 498 CGCCAGCGCGGCTACCTGAACGCCAACGTGGCTTCAACAAGTGGGCGACCGGCCCAT 557  
QY 228 CGGTGAGGACAGAGGTAAAGTTCATCGTACCCAGACGAGAACATCGTTACAACTCCAT 287  
Db 558 CTACGACGTGAAGCCCGACACAGCTTCGTGTGTAACAACGACACCATCGCCACCATCAC 617  
QY 288 GTAGAGTCTCTTCTCCATCTCTTCTGGATCAGAAATCAACAAGTGGGTCTCCAACTTGC 347  
Db 618 CGCCAGTGAATTCACCGCCCTGAACATCAAGCCCGGCGAGAGCTACCCCAAGAGGG 677  
QY 348 AGGTTACACCATCATCGACTCCGCTCAAGAACAACTCCGGTTGTCATCGGTATCATCTC 407  
Db 678 CCAGAACGSCATCGCCATCACCAGCATGAGGAGCTTCAACAGCCACCCCATCACCTGAA 737  
QY 408 CAAGTCTCTGCTTCAAGTGAAGATCTACATCAACGGTAAAGTGTGAGAGTCCATCACTTCT 467  
Db 738 CAAGAGCAGGTGGACAACTGCTGAACAAAGCCAGCATCATCGTGGAGACCAACAGAG 797  
QY 468 CTAGCAGATCTCCAAACAAAGCTCTGTTACAAACAGTGGTCTTCTCGTCAAGTCAACAA 527  
Db 798 CGAGGCGCTTCAAGATCAAGGACACCGCCACATCGTCAAGCGGCGAGTGGNA 857  
QY 528 CAACATGATGGGTAAAGATCTACATCAACGGTAAAGTGTGAGAGTCCATCACTCAAA--- 584  
Db 858 CGGCGTATCCAGCAGATCAAGGCCAAGACCGCCAGCATCATCGTGGAGCGGGGAGG 917  
QY 585 ---GGTCAAGAGTTCACCGGTATCACTTCCCAAGACCATCACTCTCGATGAGTCAACAA 641  
Db 918 CGTGCGGAGAGCGGCTGGCCCGCAGGACTACGAAACCCCGAGGACCAAGACCCCGAG 977  
QY 642 GATCCAGACACCGGCTGTGATCACTCCGACTCCGAGCAACATCAACATGTGATCGGTGA 701

Db 978 CCTGACCTGAAGAGGCGCTGAAGCTGAGCTACCCGACGAGATCAAGGAGTCGAGG 1037  
QY 702 CTTCTATATCTTCCGCAAGGATTTGGAGGTAAGGACATCAACATCTCTGTTCAACTCCTT 761  
Db 1038 CCGTCTGTACTACAAGAACAGCCCATCTACGAGAGCAGCGTGATGACCTATCTAGACGA 1097  
QY 762 CGAGTACACCAACGTCGTCAGGACTATCTGGGTAAGCACTTGATGATACACAGGAGTA 821  
Db 1098 GAACACCGGCAAGGAGGTGACCAAGCAGCTGAACGACACACCACCGGCAAGTTCAGGAGCT 1157  
QY 822 CTACATGGTCAACATCGACTACTTGAACAGATATAGTACGCCCACTCCAGACAGATCGT 881  
Db 1158 GAGCCACCTGTAGAGCTGAAGCTCAACCCCAAGATGAAGCTGACCATCAAGCTGACAT 1217  
QY 882 CTTCAACACCAAGCTATACCAACAGCTTTCAAGAGGTTTACAGAGTATCATCAAGAG 941  
Db 1218 CCGTGTACGACACCGCGAGAGCAAGCAAGCAGTCCGCAAGTGGACCAACCAACAT 1277  
QY 942 TATCAGAGGTAAACACCAAGCAGCAGAGTCAAGAGTGGTGACATCTCTGTTCTGACAT 1001  
Db 1278 CGTGAGCGGCGCACACACCGGCAAGCAGTACAGCAGCAACACCCGCGCCACCT 1337  
QY 1002 GACTATCAACACAGGCGTACACCTTTTCATGAAGAACGAGACCATGTACGCCGACAA 1061  
Db 1338 GACCCCTGAACCGCGAGCGGCGAGGAGGAGT-----GAACCAAGACCGCGACTACTAT 1391  
QY 1062 CCACCTCCACCGAGGACATCTACGCCATCGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1121  
Db 1392 CAGCCTGTACATGAGAGCGGAGAGCAACACCGAGTCCGATGATCTGTTCCATCGGTAC 1451  
QY 1122 CAACATCATCTTCCAGATCCAGCCAAATGAACAACTTACTACTACGTTCCCATGATCTT 1181  
Db 1452 ATACCCCATCAACCAAGCAGCGTGAACGTTGAACAAAGCAACTACAGGCGCTGGACAT 1511  
QY 1182 CAGTCCCACTCAACGCTGAGCAACATCTCGGTATCTGTTCCATCGGTACCTAC 1236  
Db 1512 CATGCCCAACATCAAGAGCAACCCCATCAGAGCGCTGCACATCAAGACCAAC 1566

RESULT 15  
US-08-471-044-18  
Sequence 18, Application US/08471044  
Patent No. 5840868  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSER: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,044  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2004 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..2004  
OTHER INFORMATION: /note= "Maize optimized DNA  
OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"  
US-08-471-044-18  
Query Match 6.8%; Score 93.4; DB 2; Length 2004;  
Best Local Similarity 44.3%; Pred. No. 6.1e-12;  
Matches 529; Conservative 0; Mismatches 651; Indels 15; Gaps 3;  
QY 51 CCGTGTGAAGGACATCAACAGAGTACTTCAACAACATCAACGACTCCCAAGATCTCTGTC 110  
Db 378 CCGTGAAGCGCGCATCGGTCCCAAGGCGATCAGCTTGGCGTGAGCGTGAATACCAAGCA 437  
QY 111 CCGTGAGAACCGTAAGAACACCTTGGTGTGACACCTCCGCTTACACCGCCGAGGTCTCCGA 170  
Db 438 CAGCAGACCGGTGGCGCCAGAGTGGGCGACCGACCGCGGCAACACAGCAGTCAACAC 497  
QY 171 GGAGGTGACGTCCAGCTGAACCAATCTTCCATTGCACTTCAAG---CTGGTTCCTC 227  
Db 498 CGCCAGCGCGCGCTACCTGAACGCAACGTGCGCTACAACAACGTGGCGCACCGCGCCAT 557  
QY 228 CCGTGAGGACAGAGTAAAGTCAATGCTACCCAGACAGAGACATGCTCTACACTCCAT 287  
Db 558 CTACGAGCTGAAGCCCGACCGATCGTGTCTGATGATCAGATCAACAAGTGGGTCTCCACTTGC 347  
QY 288 GTACGAGTCTTCTCTCATCTCTTCTGATCAGATCAACAAGTGGGTCTCCACTTGC 347  
Db 618 CGCCAAGTCGATTCACCGCCCTGACATCAGCCCGCGGAGAGCTACCCCAAGAGGG 677  
QY 348 AGGTACACCATCATCATGCTCCGCTCAGAACAACTCCGTTGGTTCATCGGTATCATCTC 407  
Db 678 CCAGAAGCGCATCGCCCATCACCAGCATGGAGCTTCAACAGCCACCCCATCCCTGAA 737  
QY 408 CAACCTCTCTGGTCTTCAACCTGAGCAGAGGAGTCCGAGCAGTCCCACTTCTC 467  
Db 738 CAAAGACAGGTGGAGCAACCTGCTGACAAACAGCCCATGATGCTGGAGACCAACAGAC 797  
QY 468 CTACGACATCTCCAAACACAGCTCTCTGTTTACAAACAGTGGTCTTCTGTCACCGTACCAA 527  
Db 798 CGACGCGCTCTACAAGATCAAGGACACCCACCGGCAACATGCTGACCGCGCGAGTGGAA 857  
QY 528 CAACATGATGGGTATCATGAGATCTTACATCAACGGTAAAGTGTATCGACACCATCAA--- 584  
Db 858 CCGCGTGTATCCAGAGATCAAGGCCAGACCCGCGGAGCATCATCTGTCGACGACGCGGAGCG 917



QY 585 ---GGTCAAGAGGATGACCGGTATCAACCTTCCCAAGACCATCCTTCGAGATCAACAA 641  
 Db |||||  
 QY 918 CGTGGCGGAGAGCGGTGGCCCAAGAGCTACGAGAACCCCGAGGACAGACCCCCAG 977  
 Db |||||  
 QY 642 GATCCGAGACACCGGTCTGATCACTCCGACTCCGACACATCAACATGTGGATCCGTGA 701  
 Db |||||  
 QY 978 CTTGACCTGAAGAGCGCCTGAGCTGAGCTACCCGACGAGATCAAGGAGATCGAGGG 1037  
 Db |||||  
 QY 702 CTTTATCATCTTCGCCAAGGAGTTGACCGGTAAAGGACATCAACATCCTGTTCAACTCCTT 761  
 Db |||||  
 QY 1038 CCGCTGTACTACAAAGAACAGCCCATCTACGAGACAGCGGTGATGACCTATCTAGACGA 1097  
 Db |||||  
 QY 762 GCAGTACCAACAGCTCGTCAAGAGCTACTTGGGGTACGACCTGAGATACACAAAGGAGTA 821  
 Db |||||  
 QY 1098 GAACCGCCCAAGGAGGTGACCAAGAGCTGAACGACACCCCGGCAAGTTCAAGGACGT 1157  
 Db |||||  
 QY 822 CTACATGCTCAACATGACCTTACCAAGATACATGTACGCCAACTCCAGACAGATCGT 881  
 Db |||||  
 QY 1158 GAGCCACCTGTACGACGTGAAGCTGACCCCAAGATGAACGTGACCATCAAGCTGAGCAT 1217  
 Db |||||  
 QY 882 CTTCAACACAGAGCTTAACAAACGACTTCAACGAGGTTTACAAGATCATCAAGCG 941  
 Db |||||  
 QY 1218 CCTGTACGACACCGCGAGAGCAACGACAGCATCGGCAAGTGGACCAACCAACAT 1277  
 Db |||||  
 QY 942 TATCAGAGGTAACACCAACGACACAGAGTCAAGAGGTGCTGACATCCTGTCTTGGACAT 1301  
 Db |||||  
 QY 1278 CGTACGCGCGCAACACACGCGCAAGAGCAGTACAGCAGACACCCCGACGCACT 1337  
 Db |||||  
 QY 1002 GACTATCAACAAAGCCCTACAACTGTTTCATGAGAGACGAGACCATGTACGCGGACAA 1061  
 Db |||||  
 QY 1338 GACCTGAACACCGACCGCCGAGGAGAGCT-----GAACAAGAACCCGCGACTACTACAT 1391  
 Db |||||  
 QY 1062 CCAGTCCACCGAGGACATCAACCCATCGGTCTGCTGAGCAGACCAAGGACATCAACGA 1121  
 Db |||||  
 QY 1392 CAGCTGTACATGAGAGCGAGAGACACCCAGTCCGAGATCACCATCGCGGAGAT 1451  
 Db |||||  
 QY 1122 CAACATCATCTTCAGATCCAGCCCAATGACACACATTTACTACTAGGTTCCAGATCTT 1181  
 Db |||||  
 QY 1452 ATACCCCATCACCAAGACCGCTGAACGTGAACAAGGACAACTACAGCGGCTGGACAT 1511  
 Db |||||  
 QY 1182 CAAGTCCAACTTCAACGCTGAGAACATCTCCGGTATCTGTTCCATCGGTACCTAC 1236  
 Db |||||  
 QY 1512 CATCGCCCAACATCAAGAGCAACCCCATCAGCGCCTGCAACATCAAGACCAAC 1566  
 Db |||||

Search completed: November 7, 2002, 13:46:45  
 Job time : 74 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 13:24:50 ; Search time 57 Seconds

(without alignments)  
8533.524 Million cell updates/sec

Title: US-09-910-186a-9

Perfect score: 1371

Sequence: 1 gaattcagatgacatcccc.....tctccgagtaaataggaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	1366.2	99.6	1371	10	US-09-350-756-3
2	662	48.3	3876	12	US-10-051-952-4
3	217.4	15.9	1347	10	US-09-350-756-6
4	188.2	14.5	1317	10	US-09-350-756-5
5	168.8	12.3	1338	10	US-09-350-756-1
6	162.4	11.8	1400	10	US-09-350-756-4
7	151.6	11.1	1313	8	US-08-981-087A-6
8	136.6	9.2	702	10	US-09-350-756-10
9	120.4	8.8	1341	10	US-09-350-756-2
10	93.4	6.8	1635	10	US-09-864-761-20241
11	93.4	6.8	1973	10	US-09-864-761-3471
12	92.2	6.7	1397	10	US-09-216-393-343
13	92.2	6.7	1397	10	US-09-216-393-345
14	90.2	6.6	867	10	US-09-216-393-340
15	90.2	6.6	867	10	US-09-216-393-342
16	88.8	6.5	3831	12	US-10-051-952-5
17	82.6	6.0	1959	10	US-09-864-761-4012
18	82.4	6.0	2535	10	US-09-476-242-13
19	81.6	6.0	1075	10	US-09-864-761-19241

c	20	81.6	6.0	1403	10	US-09-864-761-2513	Sequence 2513, Ap
	21	77.6	5.7	2529	10	US-09-476-242-14	Sequence 14, Appl
	22	74	5.4	2298	10	US-09-476-242-22	Sequence 22, Appl
	23	74	5.4	2298	10	US-09-476-242-23	Sequence 23, Appl
	24	74	5.4	2298	10	US-09-476-242-24	Sequence 24, Appl
c	25	73.4	5.4	583	10	US-09-864-761-20772	Sequence 20772, A
	26	71.2	5.2	2523	10	US-09-476-242-15	Sequence 15, Appl
	27	71	5.2	1818	10	US-09-801-368-427	Sequence 427, Appl
	28	70.8	5.2	2310	10	US-09-476-242-21	Sequence 21, Appl
	29	69.8	5.1	1977	12	US-10-090-624-11	Sequence 11, Appl
	30	69.8	5.1	2541	10	US-09-476-242-12	Sequence 12, Appl
	31	69.4	5.1	2010	12	US-10-032-717-9	Sequence 9, Appl
	32	68.2	5.0	2541	10	US-09-476-242-11	Sequence 11, Appl
	33	67.8	4.9	1518	9	US-09-934-060A-23	Sequence 23, Appl
	34	67.8	4.9	1518	9	US-09-934-060A-29	Sequence 29, Appl
	35	67.8	4.9	1668	9	US-09-934-060A-5	Sequence 5, Appl
	36	67.8	4.9	2159	9	US-09-934-060A-1	Sequence 1, Appl
	37	67.8	4.9	2159	9	US-09-934-060A-3	Sequence 3, Appl
	38	67.6	4.9	1944	10	US-09-864-761-2825	Sequence 2825, Ap
	39	67.6	4.9	2316	10	US-09-476-242-4	Sequence 4, Appl
	40	67.4	4.9	2517	10	US-09-476-242-16	Sequence 16, Appl
	41	67.4	4.9	2517	10	US-09-476-242-17	Sequence 17, Appl
	42	67.4	4.9	2541	10	US-09-476-242-9	Sequence 9, Appl
	43	67.4	4.9	2541	10	US-09-476-242-10	Sequence 10, Appl
	44	66.2	4.8	766	10	US-09-864-761-19608	Sequence 19608, A
	45	66	4.8	2316	10	US-09-476-242-8	Sequence 8, Appl

## ALIGNMENTS

### RESULT 1

US-09-350-756-3  
; Sequence 3, Application US/09350756  
; Patent No. US2002034521A1  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
; APPLICANT: John S. Lee  
; APPLICANT: Peter Pushko  
; APPLICANT: Michael D. Parker  
; APPLICANT: Jonathan F. Smith  
; APPLICANT: Mark T. Dertzbaugh  
; APPLICANT: Leonard Smith  
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
; FILE REFERENCE: 003/124/SAP RIID 98-21  
; CURRENT APPLICATION NUMBER: US/09/350,756  
; CURRENT FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: US 60/092,416  
; EARLIER FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
; SEQ ID NO 3  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"  
US-09-350-756-3

Query Match	99.6%	Score 1366.2;	DB 10;	Length 1371;
Best Local Similarity	99.8%	Pred. No. 0;		
Matches 1368;	Conservative	0;	Mismatches	3;
			Indels	0;
			Gaps	0;
Qy	1	GAATTCAGATGACATCCCATTCACATCTCTCTCTACACCAACATCCCTGTTGAAG	60	
Db	1	GAATTCAGATGACATCCCATTCACATCTCTCTCTACACCAACATCCCTGTTGAAG	60	
Qy	61	GACATCATCAACGAGTACTTCAACAACATCAAGACTTCCAAAGATCTGCTCGTGCAGAAC	120	
Db	61	GACATCATCAACGAGTACTTCAACAACATCAAGACTTCCAAAGATCTGCTCGTGCAGAAC	120	
Qy	121	CGTAGACACCTTGGTCGACACTCCGGTTTCAAGCCGAGGTCTCCCGAGNGGTGAC	180	

Db 121 CGTAAGAACACCTTGTGTGACACCTCCCGGTTTACACGGCCGGGTCTCCGAGAGGGTGAAC 180  
 QY 181 GTCCAGTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTTCCTCGGTGAGGACAGA 240  
 Db 181 GTCCAGTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTTCCTCGGTGAGGACAGA 240  
 QY 241 GGTAAAGTCACTGTACCCAGACAGACATCGTCTACAACTCCATGTAGAGTCCCTTC 300  
 Db 241 GGTAAAGTCACTGTACCCAGACAGACATCGTCTACAACTCCATGTAGAGTCCCTTC 300  
 QY 301 TCCATCTCTCTTGTGATGAGATACAAAGTGGGTCTCCAACTTCCAGGTACACCATC 360  
 Db 301 TCCATCTCTCTTGTGATGAGATACAAAGTGGGTCTCCAACTTCCAGGTACACCATC 360  
 QY 361 ATGAGTCCGTGAAGAACAACTCCGGTGGTCCATCGGTATCATCTCAACTTCCCTGGTC 420  
 Db 361 ATGAGTCCGTGAAGAACAACTCCGGTGGTCCATCGGTATCATCTCAACTTCCCTGGTC 420  
 QY 421 TTCACCTGAAGCAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTAGACATCTCC 480  
 Db 421 TTCACCTGAAGCAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTAGACATCTCC 480  
 QY 481 AACAAAGCTCTCGTGTACAAAGAGTGGTCTTCTGTCACCGTCCACCAACATATGGGT 540  
 Db 481 AACAAAGCTCTCGTGTACAAAGAGTGGTCTTCTGTCACCGTCCACCAACATATGGGT 540  
 QY 541 AACATGAAGATCTACATCAACGGTGAAGTGTGACACCATCAAGTCAAGGTTGACC 600  
 Db 541 AACATGAAGATCTACATCAACGGTGAAGTGTGACACCATCAAGTCAAGGTTGACC 600  
 QY 601 GGTATCAACTTCTCCAGACCAATACCTTCGAGATCAACAGATCCCAAGACCGGTCTG 660  
 Db 601 GGTATCAACTTCTCCAGACCAATACCTTCGAGATCAACAGATCCCAAGACCGGTCTG 660  
 QY 661 ATCACTTCCGACTCCGACACATCAACATGTGGTCCGTGACTTCTACATCTTCCGCCAG 720  
 Db 661 ATCACTTCCGACTCCGACACATCAACATGTGGTCCGTGACTTCTACATCTTCCGCCAG 720  
 QY 721 GAGTTGACCGGTGAAGGACATCAACATTCCTGTTCAACTCTTCCAGTACCAACAGTCTG 780  
 Db 721 GAGTTGACCGGTGAAGGACATCAACATTCCTGTTCAACTCTTCCAGTACCAACAGTCTG 780  
 QY 781 AAGGACTACTGGGTGAAGGACATCAACATTCCTGTTCAACTCTTCCAGTACCAACAGTCTG 840  
 Db 781 AAGGACTACTGGGTGAAGGACATCAACATTCCTGTTCAACTCTTCCAGTACCAACAGTCTG 840  
 QY 841 TACTTGAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACCGACGTAA 900  
 Db 841 TACTTGAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACCGACGTAA 900  
 QY 901 AACACGACTTCAACGAGGTGTACAGATCATCATCAAGCGTATCAGAGGTAAACCAAC 960  
 Db 901 AACACGACTTCAACGAGGTGTACAGATCATCATCAAGCGTATCAGAGGTAAACCAAC 960  
 QY 961 GACACGAGTCTGAGAGTGTGACATCTCTGATCTTCTGATCTTCAACAAAGGCC 1020  
 Db 961 GACACGAGTCTGAGAGTGTGACATCTCTGATCTTCTGATCTTCAACAAAGGCC 1020  
 QY 1021 TACAACTGTTCATGAAGAACGAGACATGTACGGGACACCATCTCACCAGGACATC 1080  
 Db 1021 TACAACTGTTCATGAAGAACGAGACATGTACGGGACACCATCTCACCAGGACATC 1080  
 QY 1081 TACCCATCGGTCTGGTGAGACCAAGGACATCAAGGACAACTATCTTCCAGATC 1140  
 Db 1081 TACCCATCGGTCTGGTGAGACCAAGGACATCAAGGACAACTATCTTCCAGATC 1140  
 QY 1141 CAGCCATGAACAACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGT 1200  
 Db 1141 CAGCCATGAACAACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGT 1200  
 QY 1201 GAGAACATCTCCGGTATCTGTTCCATCGGTACCTACAGATCTCCGTGGTGAAGTGG 1260  
 Db 1201 GAGAACATCTCCGGTATCTGTTCCATCGGTACCTACAGATCTCCGTGGTGAAGTGG 1260

Copied from 10910186 on 05-05-2004

QY 1261 TACAGACAACTACTTGTTCCTCAACTGTCAAGCAGGTAAGTCTCTCTCTGGAG 1320  
 Db 1261 TACAGACAACTACTTGTTCCTCAACTGTCAAGCAGGTAAGTCTCTCTCTGGAG 1320  
 QY 1321 TCCACTTCCACCACTGGGGATTCCTCCAGTCTCCGAGTATAGGAATTC 1371  
 Db 1321 TCCACTTCCACCACTGGGGATTCCTCCAGTCTCCGAGTATAGGAATTC 1371

RESULT 2

US-10-051-952-4  
 ; Sequence 4, Application US/10051952  
 ; Patent No. US20020107199A1

GENERAL INFORMATION:  
 ; APPLICANT: Walker, Patricia  
 ; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
 ; FILE REFERENCE: 2933CIP  
 ; CURRENT APPLICATION NUMBER: US/10/051,952  
 ; CURRENT FILING DATE: 2002-01-17  
 ; PRIOR APPLICATION NUMBER: 09/730,237  
 ; PRIOR FILING DATE: 2000-12-05  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4

; LENGTH: 3876  
 ; TYPE: DNA  
 ; ORGANISM: botulinum toxin  
 US-10-051-952-4

Query Match 48.3%; Score 662; DB 12; Length 3876;  
 Best Local Similarity 68.1%; Pred. No. 5.3e-156;  
 Matches 920; Conservative 0; Mismatches 430; Indels 0; Gaps 0;

QY 13 ACCATCCCATCAACATCTTCCTACACCAACTCCCTGTTGAAGGACATCATCAAC 72  
 Db 2527 ACATPACCTTTAATATTTTTCATATCTATAATTTCTTAAAGATATATTAAT 2586  
 QY 73 GAGTACTTCAACAACTCAACGACTCCAGATCTCTCCCTGCAGAACCGTAAGAACAC 132  
 Db 2587 GAATATTTCAATAATTAATGATTCAAAAATTTTGAGCCTACAAAAACAGAAAAAT 2646  
 QY 133 TTGGTCGACACCTCCGGTTTACAAACCGGAGGTCTCCGAGGAGGTGAGTCCAGCTGAC 192  
 Db 2647 TTAGTGGATACATCAGGATTAATGCAAGTGAAGTGAAGAGGCGATGTTGAGTTAAT 2706  
 QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCCTCCGGTGAGGACAGAGGTAAAGTCA 252  
 Db 2707 CCAATATTTCCATTTGACTTTAAATTTAGTAGTCTCAGGGAGGATAGAGTAAAGTTATA 2766  
 QY 253 GTCACCCAGAACGAGACATCTCTACACCTCCATGTAGAGTCTCTCTCCATCTCCTTC 312  
 Db 2767 GTAAACCCAGAAATGAAATTTGTATATTTCTATGATGAAAGTTTATGACATAGTTT 2826  
 QY 313 TGGATCAGATCAACAAAGTGGGTCTCCAACTTCCAGGTTCACACCATCATCGACTCCGTC 372  
 Db 2827 TGGATTAGATTAATAATGAGTGAAGTAAATTTTACCTGGATATATAATTAATGATAGTT 2886  
 QY 373 AAGAACAACTCCGGTGTGTCATCGGTATCATCTCCAACTCTCTCTCTCCCTCCCTCAAG 432  
 Db 2887 AAAAATAAGTAACTAGGTGGATAGTATTTATAGTAAATTTTTTTAGTATTTTACTTTAAA 2946  
 QY 433 CAGAACGAGGACTCCGAGCAGTCCATCAACTCTCTACGACATCTCCAAACAGCTCCT 492  
 Db 2947 CAAATGAGATAGTAGAACAAAGTATAATTTTAGTTATGATATATCAATATGCTCCT 3006  
 QY 493 GGTTCACAACTGGTTCCTTCCTCCCGTCCAGTACCAACACATGATGGGTAAACATGAAGATC 552  
 Db 3007 GGATAGATAAATGTTTTTTTCTAACTGTTTAACTAACTATGATGGGAAATATGAAGATT 3066  
 QY 553 TACATCAACGGTAAAGTGTATCGACACCATCAAGGTCAGAGGTGACCGGTATCACTTC 612  
 Db 3067 TATATAATGGAATAATTAATAGTACTATAAAGTTAAGAACTAACTGGGAATTAATTTT 3126





Query Match 12.3%; Score 168.8; DB 10; Length 1338;  
Best Local Similarity 54.8%; Pred. No. 5.6e-33;  
Matches 447; Conservative 0; Mismatches 327; Indels 42; Gaps 4;

QY 64 ATCATCAACAGTACTTCAACAACTCAACAGCTCCAGATCTCTCCCTCGAGAACCGT 123  
DB 27 ACCTTCACTGAATACATCAAGAAATCAATACCTCCATCTCTGAACCTGGCTACGAA 86  
QY 124 AAGAACACCTTGGTGCACCTCCGCTTACAGCGCGGAGGTCTCCGAGAGGTGACGTC 183  
DB 87 TCCAACTCACTGATCGACCTGCTCCGCTTCCGCTTCCGCTGAGGACAGAGGT 243  
QY 184 CAGCTGAACCAATCTCCATTCGACTCAAGCTGGGTCTCCGCTGAGGACAGAGGT 243  
DB 147 AACTTCGATCCGATCGCAAGATCAAGATCCAGCTG-----TTCAATCTGGAATCTCC 200  
QY 244 AAGTCACTGCTACCCAGAACAGAGAACTCGTCTACAACTCCATGATGACGCTTCCTCC 303  
DB 201 AAATCGAAGTTATCTCGTGAAGATGCTATCGTATACAACTCTATGTACGAAAATCTCC 260  
QY 304 ATCTCTCTTGGATCAAGATCAACAACTGGGTCTCCAACTTGC-----AGGTTCAC 354  
DB 261 ACCTCTCTTGGATCGTATCCGGAATCTCTCAACTCCATCTCTCTGAACATGAATAC 320  
QY 355 ACCATCATGACTCCGCTCAAGAACTCCGCTTGGTCTCATCGGTATCATCTCCAACTTC 414  
DB 321 ACCATCATCACTGCTGGAAGAACTCTGGTGGAAAGTATCTCTGAACATACGCTGAA 380  
QY 415 CTGGTCTTCACTGAGCAGAGAGGACTCCGAGCAGTCTCAACTCTCTCTTCCACGAC 474  
DB 381 ATCATCTGGACTCTGAGGACATCAAGAACTCAACACGGGTGCTGATTAATAACTCT 440  
QY 475 ATCTCCAAACACGCTCTGCTTAAACAA-----GTGGTCTTGGTCACTCCACCAACAC 531  
DB 441 CAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTTCTGATCATCAACAAAT 500  
QY 532 ATGATGGTAACTGATGATCTACATCAACGCTGATGATGATGATGATGATGATGATGAT 591  
DB 501 CGTCTGAATCACTCAAAATCTACATCAACGCGCTGCTGATGATGATGATGATGATGAT 560  
QY 592 GAGTGAACGCTGATCACTCTCCAGACATCACTCCGCTGATGATGATGATGATGATGAT 651  
DB 561 AATCTGGTAACTCACTCCGCTCTTATTAACATCATGTTCAACTGGACGGTGTGCTGAC 620  
QY 652 ACCGCTGATCACTCCGACTCCGACTCCGACTCAACATGATGATGATGATGATGATGAT 711  
DB 621 ACTC-----ACCGCTACATCTGGATCAAAATCACTCAATCTG 656  
QY 712 TTCGCCAAGAGTGGAGTGAAGGATCAACATCACTCTTCACTCTTCCGCTGATGATGAT 771  
DB 657 TTCGACAAAGAACTGAACGAAAGAAATCAAGACCTGTACGACAAACGATCCAAATCT 716  
QY 772 AACGCTGCTAAGGACTACTGGGTAAAGGACTGATGATGATGATGATGATGATGATGAT 831  
DB 717 GGTATCTGAAAGACTTCTGGGTGACTACCTGAGTGAAGGACTGATGATGATGATGATGAT 776  
QY 832 AACATGACTACTGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 857  
DB 777 AATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 812

RESULT 6  
US-09-350-756-4  
; Sequence 4; Application US/09350756  
; Patent No. US2002034521A1  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
; APPLICANT: John S. Lee  
; APPLICANT: Peter Pushko  
; APPLICANT: Michael D. Parker  
; APPLICANT: Jonathan F. Smith  
; APPLICANT: Mark T. Dertzbaugh  
; APPLICANT: Leonard Smith

; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
; FILE REFERENCE: 003/124/SAP RLIID 98-21  
; CURRENT APPLICATION NUMBER: US/09/350,756  
; EARLIER FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: US 60/092,416  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
; SEQ ID NO 4  
; LENGTH: 1400  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"  
US-09-350-756-4

Query Match 11.8%; Score 162.4; DB 10; Length 1400;  
Best Local Similarity 54.4%; Pred. No. 2.3e-31;  
Matches 468; Conservative 0; Mismatches 341; Indels 51; Gaps 5;

QY 16 ATCCCATTCACATCTTCTCTACACCAACTCCCTGTTGAAGGACATCATCAACGAG 75  
DB 67 ATTCTTTTAAGCTTCTCTTATACGATGATAAAATTTTAAATTTCTCTTCAACAG 126  
QY 76 TACTTCAACAACTCAACGACTCCAAAGATCTCTCCCTGCAAGAACCGTAAGAACACCTTG 135  
DB 127 TTCTTCAAGAGAAATTAAGTCTTCTCCGTTTAAACATGAGATCAAGAATGATAAATAC 186  
QY 136 GTGACACCTCCGTTTACACCGCGAGGCTCCGAGGAGGTGAGTCCAGTGAACCCA 195  
DB 187 GTGACACCTCCGTTTACGACTCCAAATATCAACATTAAGGTTGAGGTGATCAAGTACCCA 246  
QY 196 ATCTTCTCCATTCGACTTCAAGCTGGTCTCTCCGTTGAGGACAGAGGTAAGTCACTGTC 255  
DB 247 ACTAACAACCAATTC-----GGTATCTACACGACAAAGTTACTGAGCTGAATC 300  
QY 256 ACCGACAGACGAGAACTGCTTACAACTCCATGATGATGATGATGATGATGATGATGAT 315  
DB 301 TCTCAAAAGACTACATTTATCTACGACAAAGTACAAAGTCTCTTATTTCTTTCTGG 360  
QY 316 ATCAGAAAT-----CAACAAGTGGTCTCCAACTTCCAGGTTACACCACTATC 363  
DB 361 GTGAGGTTCTTAACTACGACAAAGATCGTCAAGTTTAAACAGAGTACACTATCATC 420  
QY 364 GATCTCGTCA---AGAACAACTCCGTTTGGTCCATCGGTATCATCTCCAACTTCCTGTC 420  
DB 421 AACTGTATGAGAGACAACTCCGTTTGGAGGTTCTCTTAAACCAACAGAGATCTT 480  
QY 421 TTCACTGAGGACAGACGAGGACTCCGAGCAGTCCATCAACTTCTCTTACGACATCTCC 480  
DB 481 TGGACCTTGAACACAAACGAGGATTAACCAAAAGTTAGATTAACCTACGTTAACGCA 540  
QY 481 AACACGCTCTCTGTTA---CAACAAGTGGTCTTGGTCAACGCTCACCAACAACTATGATG 537  
DB 541 AACGGTATTTCTGACTACATCAACAAAGTGGATTTTGGTCACTATCACTAACGACAGATTA 600  
QY 538 GGTAACTGAGATCTACATCAACGTTAGTGTGACGACCATCAAGTCAAGGAGTTG 597  
DB 601 GGTGACTCTAAGCTTTTACATTAACGTTAACTTAAGGACCAAAAGTCCATTTTAACTTA 660  
QY 598 ACCGGTATCAACTTCTCCAAAGACCATCACTCTCGAGATCAACAAGATCCGACACCGGT 657  
DB 661 GGTAACTTCACTGTTCTGACAACTCTTATTAAGATGCTTAACTGCAAGTTACACGAG 720  
QY 658 CTGATCACTCCGACTCCGACAACTCAACATGATGATGATGATGATGATGATGATGATG 717  
DB 721 T-----ACATTTGCAATTTAGTACTTCAACATTTTTCGAC 753  
QY 718 AAGGAGTTGACGAGTGAAGGACATCAACATCTCTGTTCACTCTTCACTGATGATGATGATG 777  
DB 754 AAGGAGTTAGACGAGACGAGATTCAACTTTTATACGACGACGACCACTTAACCACTAT 813  
QY 778 GTCAAGGACTACTGGGTAAACGACCTGAGATCAACAAGGAGTACTACATGCTCAACATC 837

Db 814 TTGAAGGACTTCTGGGTAACACTACTTCTTTACGACAGGAATACTACTTATTAAACGTTG 873

QY 838 GACTACTTGAACAGATACAT 857

Db 874 TTAAGCCCAACAACTTCAT 893

RESULT 7

US-08-981-087A-6  
 ; Sequence 6, Application US/08981087A  
 ; Patent No. US20020081304A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Elmore, Michael J.  
 ; APPLICANT: Kauchline, Margaret L.  
 ; APPLICANT: Minton, Nigel P.  
 ; APPLICANT: Pasechnik, Vladimir A.  
 ; APPLICANT: Titball, Richard W.  
 ; TITLE OF INVENTION: TYPE F BOTULINUM TOXIN AND USE THEREOF  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: NIXON & VANDERHVE P.C.  
 ; STREET: 1100 No. US20020081304A1ch Glebe Rd. 8th floor  
 ; CITY: Arlington  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22201-4741  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/981,087A  
 ; FILING DATE: 27-MAY-1998  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/GB96/01409  
 ; FILING DATE: 12-JUN-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 9511909.5  
 ; FILING DATE: 12-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Crawford, Arthur R.  
 ; REGISTRATION NUMBER: 25,327  
 ; REFERENCE/DOCKET NUMBER: 124-688  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-816-4000  
 ; TELEFAX: 703-816-4100  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1313 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-981-087A-6

Query Match 11.1%; Score 151.6; DB 8; Length 1313;  
 Best Local Similarity 53.3%; Fred: No. 1.1e-28;  
 Matches 465; Conservative 0; Mismatches 359; Indels 48; Gaps 5;

QY 32 TCCTCTACACCAACTCCCTGGTTGAAGGACATCATCAACGAGTACTTCAACAACATCA 91

Db 10 TGCTTACACTACGACGAAATCTCTGATCTTCAACAACACTGTACAAAAATCA 69

QY 92 ACAGCTCCAGATCTCTGTCCTGGAGACCGGTAGAACACCTTGGTGCACACCTCCGGTT 151

Db 70 AAGACAACCTCTATCTGGACATCGCTTACGAAACAAACAAATTCGACATCTCTGGCT 129

QY 152 ACAACCGCGAGTCTCCGAGGAGGTGACGTCACGCTGAACCCCAATCTTCCCATCCGACT 211

Db 130 ATGGTTCTACATCTATACACGCTGACGCTACATCTACTCTA-----CTAACCGCA 183

QY 212 TCAGCTGGGTCTCTCCGGTGAGGACAGAGGTAAAGTTCATCGTCAACCAGACGAGACA 271

Db 184 ACCAGTTCGGTATCTACTTCTTAAACCCGCTCTGAAGTAAACATCGCTCAGAACAGACA 243

QY 272 TCGTCTACAACTCCATGTACGAGTCCCTTCCCATCTCTTCTCTGATCAGAACTCAACAAGT 331

Db 244 TCATCTACACGCTCGTTACGAGAACTTCTCTATCTCTTCTGGTTCGTATCCGAAAT 303

QY 332 GGGTCTCCAACTT-----GCCAGGTACACCAATCATCGACTCCGTC---AAGAACA 379

Db 304 ACTTCAACAAAGTTAAAGTCAACGTAACAAATACATCATCATCATCATCGATCCGTAACA 363

QY 380 ACTCGGTTGGTTCATCGGTATCATCTCCAACTTCTCCCTCCCTCAACCTCAACGACGACG 439

Db 364 ACTCTGGTTGGAAATCTCTCTGACATACACAAATCATCTGGACTCTCGACGACACTG 423

QY 440 AGGACTCCGACGAGTCCATCAACTTCTCCACGACATCTCCAAACAGGCTCTGGTTAC- 498

Db 424 CTGGTAACAACGAGAACTGGTTTCAACTACACTCAGATGATCTCTATCTCTGACTACA 483

QY 499 --AACAAGTGTCTCTGTCACCGTCAACACATCATGATGGTAAACATGAAGATCTACA 556

Db 484 TTAATAAATGGATCTCTGTTACTATCATCAACACCTCTGGTAACTCTCTGATCTTACA 543

QY 557 TCACGCTAAGCTGATGACACACCATCAAGGTCAAGAGTTGACCGGTATCAACTTCTCCA 616

Db 544 TCAACGTTAACTGATGATGATAAATACTATCTTAACCTGGGTGACATCCAGTTTCTG 603

QY 617 AGACATCACTCTGAGATCAACAGATCCAGACACCGGTCTGATCACCCTCCGACTCCG 676

Db 604 ACAACATCTCTGTTCAAAATCTGGTTGCAACGACACGCTT----- 645

QY 677 ACAACATCAACATGTGGATCGTCTGACTTCTTACATCTTCCGCAAGGAGTTGGACGTAAG 736

Db 646 -----ACGTTGGTATCGTTACTTCAAAAGTTTTCGACACTGAACTGGTAAACTG 696

QY 737 ACATCAACATCTCTGTTCAACTCTTGCAGTACACCACTGCTCGTCAAGGACTACTGGGTA 796

Db 697 AAATCGAAACTCTGTACTCTGACGACCGGACCGTCTATCTCTGAAAGACTTCTGGGTA 756

QY 797 AGGACCTGAGATACAAAGGAGTACTACATGTGTCACATCGTCAACATCGACTTCTGAACAGATACA 856

Db 757 ACTACTGCTGTACAAACAAAGTTACTTACCTGTGTAACCTGCTCCGAGACTGACAAATCTA 816

QY 857 TGTAGCCCAACTCCAGACAGATCGTCTTCAAC 888

Db 817 TCACCTCAGAACTCTAACTTCTCTGAACATCAAC 848

RESULT 8  
 US-09-350-756-10  
 ; Sequence 10, Application US/09350756  
 ; Patent No. US20020034521A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
 ; APPLICANT: John S. Lee  
 ; APPLICANT: Peter Pushko  
 ; APPLICANT: Michael D. Parker  
 ; APPLICANT: Jonathan F. Smith  
 ; APPLICANT: Mark T. Dertzbaugh  
 ; APPLICANT: Leonard Smith  
 ; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
 ; FILE REFERENCE: 003/124/SAP R1D 98-21  
 ; CURRENT APPLICATION NUMBER: US/09/350,756  
 ; CURRENT FILING DATE: 1999-07-09  
 ; EARLIER APPLICATION NUMBER: US 60/092,416  
 ; EARLIER FILING DATE: 1998-07-10  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
 ; SEQ ID NO 10  
 ; LENGTH: 702  
 ; TYPE: DNA

```
; ORGANISM: Clostridium botulinum
; FEATURE:
US-09-350-756-10

Query Match      9.2%; Score 126.6; DB 10; Length 702;
Best Local Similarity 53.6%; Pred. No. 1.5e-22;
Matches 382; Conservative 0; Mismatches 289; Indels 42; Gaps 4;

Oy 64 ATCATACAGAGTACTTCAACACATCAACAGTCTCAAGTCTCTGCTCCGTCGAGACCGT 123
Db 19 ACCTTCACTGAATACATCAAGACATCATCAATACCTCCATCTGAACCTCGCTACGAA 78

Oy 124 AAGAACACCTTGGTCGACACCTCGGTTACAACGCGAGGTCTCCGAGAGGTGACGTC 183
Db 79 TCCAATCACCTGATCGACCTGTCTGCTAGCTTCCGATTCACAAATCAACATCGTTCTAAAGTT 138

Oy 184 CAGCTGAACCAATCTTCCATTCGACTTCAAGTGGTCTCCGCTGAGGACAGAGT 243
Db 139 AACTTCGATCGGATCGACAGAAATCAGATCCAGCTG-----TTCAATCTGGAATCTTCC 192

Oy 244 AAGTCTATCGTCACCCAGAACGAGACATCGTCTACAACCTCCATGTACGAGTCTTCTCC 303
Db 193 AAAATCGAAGTTATCCGGAAGATGCTATCGTATACAACCTCTATGTACGAAAACITCTCC 252

Oy 304 ATCTCCCTTCGATCAGAAATCAACAGTGGTCTCAACTTGGC-----AGTTCAC 354
Db 253 ACCCTCTCTGATCCGATCCCGAATATCTCAACTCCATCTCTGAACAATGAATAC 312

Oy 355 ACCATATCGACTCCGTCAGAACCACTCCGCTTGGTTCATCGGTATCTCCAACTTC 414
Db 313 ACCATATCAACTCGATGGAAACAAATCTCTGTTGGAAGTATCTCTGAACACGGTGA 372

Oy 415 CTGGTCTTACCCCTGAGCAGAGAGGAGTCTCGAGAGTCCATCAACTTCTCTACGAC 474
Db 373 ATCATCTGACTCTGCGAGGACATCAGGAATCAACAGCGTCTTGTATCAAACTACT 432

Oy 475 ATCTCCAAACAGCTCTCTGTTTACAAACAA---GTGGTCTTCTGCTCCGCTCAACAAAC 531
Db 433 CAGATGATCAACATCTCTGACTACATCAATCGTGGATCTTCTGTACCATCAACACAT 492

Oy 532 ATGATGGTATACATGAGATCTACATCAACGGTAACTGATGACACCATCAAGGTCAG 591
Db 493 CGTCTGATTAACCTCAAAATCTACATCAACGGCGCTCTGATCGACGAGAAACGATCTCC 552

Oy 592 GAGTTGACCGGTATCAACTCTCCAGACCATCACTTCGAGATCAACAGATCCOAGAC 651
Db 553 AATCTGGTATACATCCAGGTTCTATATACATCATGTTCAACTGGACGGTGTCTGTAC 612

Oy 652 ACCGGTCTGATCACTCCGACTCCGACCAACATCAACATGTGGATCGGTACTTCTACATC 711
Db 613 ACTC-----ACCGCTACATCTGGATCAAAATACATCAATCTG 648

Oy 712 TTCGCCAAGGAGTTGGCGGTAGGACATCAACATCTCTGTTCAACTCCTTGA 764
Db 649 TTCGCAAGAAGTCTGACGAAAGAAAGAAATCAAAGACCTGTACGACCAACCACTGA 701

RESULT 9
US-09-350-756-2
; Sequence 2, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09

; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 2
; LENGTH: 1341
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-2

Query Match      8.8%; Score 120.4; DB 10; Length 1341;
Best Local Similarity 50.7%; Pred. No. 6.9e-21;
Matches 412; Conservative 0; Mismatches 346; Indels 54; Gaps 3;

Oy 64 ATCATCAACAGTACTTCAACACATCAACAGTCTCAAGTCTCTGCTCCGTCGAGACCGT 123
Db 10 ATGCCCAACAAATACAAATCCGAAATCCCTGAACAATATCATCTGAACCTCGGTACAAA 69

Oy 124 AAGAACACCTTGGTCGACACCTCGGTTACAACGCGAGGTCTCCGAGGAGGTGACGTC 183
Db 70 GACAACAATCTGATCGATCTGCTGTTACGGTCTAAAGTTGAAGTATACGACGCTGTT 129

Oy 184 CAGCTGAACCAATCTTCCATTCGACTTCAAGTGGTCTCCGCTGAGGACAGAGT 243
Db 130 GAAGTGAATGACAAGAACCACTTCAAACTGA-----CCTCTTCCGCTAACTCT 177

Oy 244 AAGTCTATCGTCAACCCAGAACGAGACATCGTCTACAACCTCCATGTACGAGTCTCTCC 303
Db 178 AAGATCGGTGTTACTCAGATCAGACATCATCTTCAACTCGGTATCTCTGGACTTCTCT 237

Oy 304 ATCTCCCTTCGATCAGAAATCAACAAAGTG-----GGTCTCCAACTTG 345
Db 238 GTTTCCTTCTGGATCTCGTATCCGAAATCAAGAACGAGGTATCCAGAAATACATCCAC 297

Oy 346 CCAGTTTACACCATCATCTGACTCCGTCACGAACAACCTCCGTTGGTCCCTCGTATCATC 405
Db 298 AATGAATACACCATCATCTGACTCATGATGAGAAATCACTCTGGTTGGAATCTCCATCCG 357

Oy 406 TCCAATCTCTGTTCTTCACTCCCTGAGCAGAGAGAGTCCGAGCAGTCCATCAACTTC 465
Db 358 GGTAACTGATCATCTGGACTCTGATATCATCAAGGTAAGACCAAAATCTGTATTTCTC 417

Oy 466 TCCTACGACATCTCCAAACAGCTCCCTGTTTACAACAAGTGGTCTTCGTCACCGTCAAC 525
Db 418 GAATACACATCTCCGTAAGACATCTCTGAATACATCAATCGCTGGTCTCTGTTACCATC 477

Oy 526 AACAAATGATGGTAAACATGAAGATCTACATCAACGCGTAAGCTGATCGACACCATCAAG 585
Db 478 ACCAATAACCTGAAACAATGCTAAATCTACATCAACGGTAAACTGGAATCTAATACCGAC 537

Oy 586 GTCAAGGAGTTGACCGGTATCACTTCTCAAGACCATCACTTCCGAGATCAACAGATC 645
Db 538 ATCAAGACATCTCGTGAAGTTATCGGTAAAGTAAATCATCTTCAAACTGGACGGTGCAC 597

Oy 546 CCAGACACCGGTCTGTGATCACTCCGACTCCGACACATCAACATGTGGATCGGTGACTTC 705
Db 598 ATCCATCTACCCAGTTTCACT-----TGATGAATATCTC 633

Oy 706 TACATCTTCGCCAAGGAGTTGGAGGTAGGACATCAACATCTCTGTTCACTCTTGGAG 765
Db 634 TCCATCTTCAACACCGCACTCTCTCAGTCCCAATATGAAGAACGGTACAAAGATCCAGT 693

Oy 766 TACACCAACGTCGTCAGGAGTACTTCTGGGTGAAGCTGTGATACAAACAAAGGAGTACTAC 825
Db 694 TACTCCGATACCTGGAAGACTTCTGGGTGAATCCCGCTGATGTACAAACAAAGATACTAT 753

Oy 826 ATGCTCAACTCGACTTACTTGAACAGATACAT 857
Db 754 ATGTTCAATGCTGTGTAACAAGAACTCTTACAT 785
```





Qy 1085 CCATCGTCTGCGTGCAGACACGACATCAACAGACATCACTCTTCCAGATCCAGC 1144  
Db 1227 CCACACCATCACTACCATCACCACCATCACCACCATCACTACACACCATCA 1285  
Qy 1145 CAATGAACAACACTTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAGA 1204  
Db 1287 CTACACCATCACTACCATCACCACCATCACTACACCATCACTACCATCACTT 1345  
Qy 1205 ACATCTCGGTATCTGTTCATCGGTACCTACA 1237  
Db 1347 CCACACCATCACTACCATCACCACCATCACTACCATCA 1379

RESULT 11  
US-09-864-761-3471  
Sequence 3471, Application US/09864761  
Patent No. US2002040763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 3471  
LENGTH: 1973  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC006547.9  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9  
OTHER INFORMATION: EXPRESSED IN HB100, SIGNAL = 12  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12  
US-09-864-761-3471

Query Match 6.8%; Score 93.4; DB 10; Length 1973;  
Best Local Similarity 43.3%; Pred. No. 4.6e-14;  
Matches 534; Conservative 0; Mismatches 696; Indels 3; Gaps 2;

Qy 5 TCAGATGACCATCCGATTCACATCTTCTCTACACCAACACCTCCCTGTTGAAGGACA 64  
Db 304 TCATACACCATCCACCATCTACATACACCAACACCATCTACATCTACATCCACCA 363  
Qy 65 TCATCAACAGTACTTCAACACATCAACGACTCCAGATCTCTGTCCTGTCGAGAACCTTA 124  
Db 364 CCATCAACATCACTACCATCCACCAACCATCACTACATCCACCACTACATCACTACCA 423  
Qy 125 AGACACCTTGGTCGACACTCCGGTTACACGCGAGGTCTCCGAGGAGGTGAGCTCC 184  
Db 424 CCACCAACCATCACTACCATCTACATCTACATCTACATCTACATCTACATCTACAT 482  
Qy 185 AGGTGAACCAATCTTCCCATCTCGACTTCAAGCTGGGTCTCCCTCGGTGAGGACAGGTA 244  
Db 483 ACCACATCACTACACCATCTACATCCACCAACCATCACTACATCACTACATCACTAC 540  
Qy 245 AGGTCACTGTCACGAGAGAGGAACTGTTCAACACTTCAATGTACGAGTCTCTCTCA 304  
Db 541 CCATCACTACATCACTACCATCCACCAACCATCACTACATCCACCACTACATCACTAC 600  
Qy 305 TCTCTTCTGGATCAGAACTCAACAAGTGGGTCTCCAACTTGCAGGTTACACCATCATCG 364  
Db 601 CCACCAACCATCACTACATCACTACATCACTACATCACTACATCACTACATCACTAC 560  
Qy 365 ACTCGTCAAGAACTCCGGTTGGTTCATCTGATCTCCAACTTCTCTGGTCTCA 424  
Db 661 CTACCAACCATCACTACATCACTACATCACTACATCACTACATCACTACATCACTAC 720  
Qy 425 CCCTGAAGCAGAGGAGTCTCCGAGCTGTCATCACTTCTCTACGACATCTCCAACA 484  
Db 721 CCACCAACCATCACTACATCACTACATCACTACATCACTACATCACTACATCACTAC 780  
Qy 485 AGCTCTCTGTTTACAAAGTGGTCTTTCGTCACCGTCACTACCAACATCACTACATCACT 544  
Db 781 TCACTGCCACCATCACTACATCACTACATCACTACATCACTACATCACTACATCACTAC 840  
Qy 545 TGAAGATCTACATCAAGGTAAGTGTATCGACCATCAAGGTCAAGGATTCAGCGTA 604  
Db 841 CCACCAACCATCACTACATCACTACATCACTACATCACTACATCACTACATCACTAC 900  
Qy 605 TCAACTTCTCAAGACCATCACTTCGAGATCAACAGATCCAGACACCGGTCTGATCA 664  
Db 901 CCACCAACCATCACTACATCACTACATCACTACATCACTACATCACTACATCACTAC 960  
Qy 565 CTTCCGACTCGGACATCAACATGTGGTTCGTCATCTTCTACATCTTTCGCGCAAGGAT 724  
Db 961 CCACCAACCATCACTACATCACTACATCACTACATCACTACATCACTACATCACTAC 1020  
Qy 725 TGGACGGTAAGGATCAACATCTCTTCAACTCTTTCAGTACACCAACCTGCTCAAGG 784  
Db 1021 CCACATCACTACATCACTACATCACTACATCACTACATCACTACATCACTACATCA 1080  
Qy 785 ACTACTGGGTACGACCTGAGATCAACAGGAGTACTACATGGTCACTACATCTGACTACT 844  
Db 1081 CCACCAACCATCACTACATCACTACATCACTACATCACTACATCACTACATCACTAC 1140  
Qy 845 TGAACAGATACATGACGCAACTCCAGACAGATCTCTTCAACACCAACCTGCTCAAGG 904  
Db 1141 CCACATCACTACATCACTACATCACTACATCACTACATCACTACATCACTACATCA 1200

QY 905 AGACATTCACGAGGTTTACAGATCATCATCAAGCGTATCAGAGGTAAACACCAACGACA 964  
 Db 1201 CCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1260  
 QY 965 CCAGAGTCAGAGGTGGTGCATCTCTGTTCCGATGATGATGATGATGATGATGATGATGAT 1024  
 Db 1261 CCATCATCACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1320  
 QY 1025 ACCTGTTTATGAGACGAGACCATGTTAGCGGACCAACCAACCAACCAACCAACCAACCAAC 1084  
 Db 1321 CCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1380  
 QY 1085 CCATCGGTGCTGGTGCAGAGACGAGGACATCAACGACATCATCTTCCAGATCCAGC 1144  
 Db 1381 CCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1440  
 QY 1145 CAATGAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1204  
 Db 1441 CTACCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1500  
 QY 1205 ACATCTCCGGTATCTGTTCCATCGGTACTACA 1237  
 Db 1501 CCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1533

RESULT 12  
 US-09-216-393-343  
 Sequence 343, Application US/09216393  
 Patent No. US2001001447A1  
 GENERAL INFORMATION:  
 APPLICANT: Milhausen, Michael James  
 TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND  
 FILE REFERENCE: TX-1-C2  
 CURRENT APPLICATION NUMBER: US/09/216,393  
 EARLIER FILING DATE: 1998-12-18  
 EARLIER FILING DATE: 1997-12-19  
 NUMBER OF SEQ ID NOS: 364  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 343  
 LENGTH: 1397  
 TYPE: DNA  
 ORGANISM: Toxoplasma gondii  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (238)..(1104)  
 US-09-216-393-343

Query Match 6.7%; Score 92.2; DB 10; Length 1397;  
 Best Local Similarity 51.1%; Pred. No. 7.9e-14;  
 Matches 217; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 734 AGGACATCAACATCTGTTCAACTCTTCCAGTACACCAACGTCGTAAGGACTACTGGG 793  
 Db 688 AGTTCACACAGTACGAGTACAGCCAGCAGCATCACTACTACTACTACTACTACTACTACT 747  
 QY 794 GTACGACCTGAGATACACAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 853  
 Db 748 ACCACTACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 807  
 QY 854 ACATGTCGCACTCCAGACAGATGCTTTCACACGACGAGGTAAACCAACGACTTCA 913  
 Db 808 ACACCAACACACGACACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 867  
 QY 914 ACAGGGTTTACAGATCATCATCAAGCGTATCAGAGGTAAACCAACCAACCAACCAACCA 973  
 Db 868 ACAACTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 927  
 QY 974 GAGGTGGTGCATCTGTTCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 1033  
 Db 928 ACCACTACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 987

QY 1034 TGAAGAAGAGAGACCATGTACCGGACAAACCAACCAACCAACCAACCAACCAACCAACCA 1093  
 Db 988 ACNACTTACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1047  
 QY 1094 TGGGTGAGAGAGAGGACATCAACGACATCATCTTCCAGATCCAGCCATATGAACA 1153  
 Db 1048 ACTACACGACTTACGACACCATCTACGAGACATCTACGAGACATCTACGACCATCTACCTCGATGATAGCC 1107  
 QY 1154 ACACT 1158  
 Db 1108 AAAGT 1112

RESULT 13  
 US-09-216-393-345/c  
 Sequence 345, Application US/09216393  
 Patent No. US2001001447A1  
 GENERAL INFORMATION:  
 APPLICANT: Milhausen, Michael James  
 TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND  
 FILE REFERENCE: TX-1-C2  
 CURRENT APPLICATION NUMBER: US/09/216,393  
 EARLIER FILING DATE: 1998-12-18  
 EARLIER FILING DATE: 1997-12-19  
 NUMBER OF SEQ ID NOS: 364  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 345  
 LENGTH: 1397  
 TYPE: DNA  
 ORGANISM: Toxoplasma gondii  
 US-09-216-393-345

Query Match 6.7%; Score 92.2; DB 10; Length 1397;  
 Best Local Similarity 51.1%; Pred. No. 7.9e-14;  
 Matches 217; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 734 AGGACATCAACATCTGTTCAACTCTTCCAGTACACCAACGTCGTAAGGACTACTGGG 793  
 Db 710 AGTTCACACAGTACGAGTACAGCCAGCAGCATCACTACTACTACTACTACTACTACTACT 651  
 QY 794 GTACGACCTGAGATACACAGGAGTACTACTACTACTACTACTACTACTACTACTACTACT 853  
 Db 650 ACACCTACGACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 591  
 QY 854 ACATGTACGCCCAACTCCAGACAGATGCTTTCACACCAACGAGGTAAACCAACGACTTCA 913  
 Db 590 ACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 531  
 QY 914 ACAGGGTTTACAGATCATCATCAAGCGTATCAGAGGTAAACCAACCAACCAACCAACCA 973  
 Db 530 ACAACTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 471  
 QY 974 GAGGTGGTGCATCTGTTCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 1033  
 Db 470 ACCACTACGACTTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 411  
 QY 1034 TGAAGAAGAGAGACCATGTACCGGACAAACCAACCAACCAACCAACCAACCAACCAACCA 1093  
 Db 410 ACAACTTACACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 351  
 QY 1094 TGGGTGAGAGAGGACATCAACGACATCATCTTCCAGATCCAGCCATATGAACA 1153  
 Db 350 ACTACACGACTTACGACACCATCTACGAGACATCTACGAGACATCTACGACCATCTACCTCGATGATAGCC 291  
 QY 1154 ACACT 1158  
 Db 290 AAAGT 286

RESULT 14

```

US-09-216-393-340
; Sequence 340, Application US/09216393
; Patent No. US2001001447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 340
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(867)
; S-09-216-393-340

Query Match      6.6%; Score 90.2; DB 10; Length 867;
Best Local Similarity 51.1%; Pred. No. 2e-13;
Matches 212; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY 734 AGGACATCAACATCCTGTTCAACTCCTTGTCAGTACACCAACGTCGTCAGGACTACTGGG 793
Db 451 AGTTATCCACAACTAGCAGTACAAACGACGACATCAACTACCACTACTACTACCACTACC 510
QY 794 GTAACGACCTGAGATACAAAGAGGAGTACTACATGGTCAACTCGACTACTTGAACAGAT 853
Db 511 ACCACTAGGACTACTACTACAACTACGACACCAACAACTACAAACACCAACCAACT 570
QY 854 ACATGTAGCCCACTCCAGACAGATCGTCTTCAACAGCAGACGTCATCAACAGAGTTCA 913
Db 571 ACACCAACAAACGACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 630
QY 914 ACGAGGTTTCAAGATCATCATCAAGCGTATCAGAGGTATACCAACCAACCAACCAAGTCA 973
Db 631 ACAACTACCAACAAACGACAAACCAACCAACCAACCAACCAACCAACCAACCAACCA 690
QY 974 GAGGTGGTGCATCCTGTACTTGCAGATGACTATCAACAAAGGCTTCAACAGCTTTCA 1033
Db 691 ACCACTAGGACTACCAACAACCTACTACTACCAACTACCAACTACCAACTACCAACTACCA 750
QY 1034 TCAGAACGAGACCATGTAGCGCGACAAACCACTCCACCGAGGACATCTAGCCATCGGTC 1093
Db 751 ACAACTACCAACAAACGACAAACCAACCAACCAACCAACCAACCAACCAACCAACCA 810
QY 1094 TCGGTGACGACGACCAAGGACATCAACGACAAACATCATCTTCCAGATCCAGCCCAAT 1148
Db 811 ACTACAAGGACTACGACACCATCTACGACGACATCCACCACCACTACCTCCCTCGATT 865

Search completed: November 7, 2002, 14:41:50
Job time : 68 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:21:30 ; Search time 3155 Seconds  
(without alignments)  
10925.626 Million cell updates/sec

Title: US-09-910-186A-9  
Perfect score: 1371  
Sequence: 1 gaattcacgagaccatccc.....tctccgagtaaggaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main.\*

- 1: /cgn2\_6/ptodata/1/pna/pctus\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq.\*
- 8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq.\*
- 17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq.\*
- 18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq.\*
- 19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*
- 20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*
- 21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*
- 22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*
- 23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq.\*
- 24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq.\*
- 25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq.\*
- 26: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq.\*
- 27: /cgn2\_6/ptodata/1/pna/US096F\_COMB.seq.\*
- 28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq.\*
- 29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq.\*
- 30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq.\*
- 31: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq.\*
- 32: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq.\*
- 33: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq.\*
- 34: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq.\*
- 35: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq.\*
- 36: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq.\*
- 37: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq.\*
- 38: /cgn2\_6/ptodata/1/pna/US100A\_COMB.seq.\*
- 39: /cgn2\_6/ptodata/1/pna/US100B\_COMB.seq.\*
- 40: /cgn2\_6/ptodata/1/pna/US101A\_COMB.seq.\*
- 41: /cgn2\_6/ptodata/1/pna/US101B\_COMB.seq.\*
- 42: /cgn2\_6/ptodata/1/pna/US102A\_COMB.seq.\*
- 43: /cgn2\_6/ptodata/1/pna/US102B\_COMB.seq.\*

- 44: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq.\*
- 45: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq.\*
- 46: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq.\*
- 47: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq.\*
- 48: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq.\*
- 49: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq.\*
- 50: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq.\*
- 51: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq.\*
- 52: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq.\*
- 53: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq.\*
- 54: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq.\*
- 55: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq.\*
- 56: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq.\*
- 57: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq.\*
- 58: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq.\*
- 59: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq.\*
- 60: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq.\*
- 61: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq.\*
- 62: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq.\*
- 63: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq.\*
- 64: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq.\*
- 65: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq.\*
- 66: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq.\*
- 67: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq.\*
- 68: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq.\*
- 69: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq.\*
- 70: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq.\*
- 71: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq.\*
- 72: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq.\*
- 73: /cgn2\_6/ptodata/1/pna/US6029\_COMB.seq.\*
- 74: /cgn2\_6/ptodata/1/pna/US6030\_COMB.seq.\*
- 75: /cgn2\_6/ptodata/1/pna/US6031\_COMB.seq.\*
- 76: /cgn2\_6/ptodata/1/pna/US6032\_COMB.seq.\*
- 77: /cgn2\_6/ptodata/1/pna/US6033\_COMB.seq.\*
- 78: /cgn2\_6/ptodata/1/pna/US6034\_COMB.seq.\*
- 79: /cgn2\_6/ptodata/1/pna/US6035\_COMB.seq.\*
- 80: /cgn2\_6/ptodata/1/pna/US6036\_COMB.seq.\*
- 81: /cgn2\_6/ptodata/1/pna/US6037\_COMB.seq.\*
- 82: /cgn2\_6/ptodata/1/pna/US6038\_COMB.seq.\*
- 83: /cgn2\_6/ptodata/1/pna/US6039\_COMB.seq.\*
- 84: /cgn2\_6/ptodata/1/pna/US6040\_COMB.seq.\*
- 85: /cgn2\_6/ptodata/1/pna/US6041\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1371	100.0	1371	23	US-09-611-419A-9
2	1371	100.0	1371	34	US-09-910-186A-9
3	1366.2	99.6	1371	17	US-09-350-756-3
4	665.2	48.5	3876	1	PCT-US97-15394-59
5	665.2	48.5	3876	11	US-08-704-159-59
6	665.2	48.5	3876	42	US-10-205-516-5
7	665.2	48.5	3950	13	US-08-954-302-2
8	665.2	48.5	3906	42	US-10-205-516-19
9	663.2	48.4	3876	29	US-09-730-237-4
10	662	48.3	3876	39	US-10-051-952-4
11	662	48.3	3876	39	PCT-US97-15394-61
12	649.8	47.4	1502	11	US-08-704-159-61
13	649.8	47.4	1502	43	US-10-271-012-61
14	649.8	47.4	1502	43	US-09-611-419A-11
15	420.2	30.6	1374	23	US-09-910-186A-11
16	417	30.4	1374	34	US-09-910-186A-11
17	217.4	15.9	1347	17	US-09-350-756-6
18	217.4	15.9	1368	23	US-09-611-419A-17
19	217.4	15.9	1368	34	US-09-910-186A-17
20	199.8	14.6	1317	23	US-09-611-419A-15
21	199.8	14.6	1317	34	US-09-910-186A-15

22 198.2 14.5 1317 17 US-09-350-756-5  
 23 169.6 12.4 1351 1 PCT-US97-15394-35  
 24 169.6 12.4 1351 11 US-08-704-159-35  
 25 169.6 12.4 1351 43 US-10-271-012-35  
 26 169.4 12.4 1323 23 US-09-611-419A-3  
 27 169.4 12.4 1323 34 US-09-910-186A-3  
 28 169.4 12.4 1402 1 PCT-US97-15394-25  
 29 169.4 12.4 1402 8 US-08-405-496-25  
 30 169.4 12.4 1402 8 US-08-422-711-25  
 31 169.4 12.4 1402 8 US-08-480-604-25  
 32 169.4 12.4 1402 11 US-08-704-159-25  
 33 169.4 12.4 1402 14 US-09-084-517-25  
 34 169.4 12.4 1402 43 US-10-271-012-25  
 35 168.8 12.3 1326 23 US-09-611-419A-5  
 36 168.8 12.3 1326 34 US-09-910-186A-5  
 37 168.8 12.3 1330 1 PCT-US97-15394-22  
 38 168.8 12.3 1330 8 US-08-405-496-22  
 39 168.8 12.3 1330 8 US-08-422-711-22  
 40 168.8 12.3 1330 8 US-08-480-604-22  
 41 168.8 12.3 1330 11 US-08-704-159-22  
 42 168.8 12.3 1330 14 US-09-084-517-22  
 43 168.8 12.3 1330 43 US-10-271-012-22  
 44 168.8 12.3 1332 23 US-09-611-419A-1  
 45 168.8 12.3 1332 34 US-09-910-186A-1

ALIGNMENTS

RESULT 1  
 US-09-611-419A-9  
 Sequence 9, Application US/09611419A  
 GENERAL INFORMATION:  
 APPLICANT: Smith, Leonard A.  
 APPLICANT: Byrne, Michael P.  
 APPLICANT: Middlebrook, John L.  
 APPLICANT: Lapenotiere, Hugh  
 APPLICANT: Clayton, Michael A.  
 APPLICANT: Brown, Douglas R.  
 TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM  
 TITLE OF INVENTION: NEUROTOXIN  
 FILE REFERENCE: A33626 067322.0105  
 CURRENT APPLICATION NUMBER: US/09/611.419A  
 CURRENT FILING DATE: 2000-07-06  
 PRIOR APPLICATION NUMBER: PCT/US00/12890  
 PRIOR FILING DATE: 2000-05-12  
 PRIOR APPLICATION NUMBER: 60/133,865  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,866  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,867  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,868  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,869  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/146,192  
 PRIOR FILING DATE: 1999-07-29  
 NUMBER OF SEQ ID NOS: 42  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 9  
 LENGTH: 1371  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Synthetic construct based on BONTA HC  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (10)...(1359)  
 US-09-611-419A-9

Query Match 100.0% Score 1371; DB 23; Length 1371;  
 Best Local Similarity 100.0% Pred. No. 3.8e-279;

	Matches	1371;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	GAATTCAGATGACATCCCATTCACATCTCTCTACACCAACACTCCCTGTTGAAG	60							
Db	1	GAATTCAGATGACATCCCATTCACATCTCTCTACACCAACACTCCCTGTTGAAG	60							
Qy	61	GACATCATCAACAGAGTACTTCAACAACATCAAGACTTCCAGATCTCTGTCGAGAAC	120							
Db	61	GACATCATCAACAGAGTACTTCAACAACATCAAGACTTCCAGATCTCTGTCGAGAAC	120							
Qy	121	CGTAAAGAACACCTTGGTCGACACCTCCGGTTACACCGCCAGGCTCCCGAGGAGGTGAC	180							
Db	121	CGTAAAGAACACCTTGGTCGACACCTCCGGTTACACCGCCAGGCTCCCGAGGAGGTGAC	180							
Qy	181	GTCCAGCTGAACCCCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCCGGTGAGGACAGA	240							
Db	181	GTCCAGCTGAACCCCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCCGGTGAGGACAGA	240							
Qy	241	GGTAAAGTCTATCGTACCCAGAGAGAAATCGTCTACAACTCCATGTACGAGTCTCTTC	300							
Db	241	GGTAAAGTCTATCGTACCCAGAGAGAAATCGTCTACAACTCCATGTACGAGTCTCTTC	300							
Qy	301	TCCATCTCTCTGATCAGAAATCAACAAGTGGGTCTCCAACTTGGCAGGTTACACCATC	360							
Db	301	TCCATCTCTCTGATCAGAAATCAACAAGTGGGTCTCCAACTTGGCAGGTTACACCATC	360							
Qy	361	ATCGACTCCGTCAAGAAACACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCCCTGGTC	420							
Db	361	ATCGACTCCGTCAAGAAACACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCCCTGGTC	420							
Qy	421	TTCCACCTGAAGCAGAGAGGACTCCGAGCAGTCCACTCAACTTCTCTACGACATCTCC	480							
Db	421	TTCCACCTGAAGCAGAGAGGACTCCGAGCAGTCCACTCAACTTCTCTACGACATCTCC	480							
Qy	481	AACAAGCTCTCTGTACAAAGTGGTCTTCTGTCACCGTCAACCAACATCATGTATGGGT	540							
Db	481	AACAAGCTCTCTGTACAAAGTGGTCTTCTGTCACCGTCAACCAACATCATGTATGGGT	540							
Qy	541	AACATGAAGATCTACATCAACCGTAAAGTGTATGACACCATCAAGTCAAGGAGTGAAC	600							
Db	541	AACATGAAGATCTACATCAACCGTAAAGTGTATGACACCATCAAGTCAAGGAGTGAAC	600							
Qy	601	GGTATCAACTCTCCAGAACATCACTTTCGAGATCAACAGATCCAGACACCGGTCGTC	660							
Db	601	GGTATCAACTCTCCAGAACATCACTTTCGAGATCAACAGATCCAGACACCGGTCGTC	660							
Qy	661	ATCAGCTCCGACTCCGACCAATCAACATGATGATCGGTCTCTACATCTTCCGCCAAG	720							
Db	661	ATCAGCTCCGACTCCGACCAATCAACATGATGATCGGTCTCTACATCTTCCGCCAAG	720							
Qy	721	GAGTTGACGCTAAGGACATCAACATCTCTCAACTCCCTTTCAGTACACCAACGTCGTC	780							
Db	721	GAGTTGACGCTAAGGACATCAACATCTCTCAACTCCCTTTCAGTACACCAACGTCGTC	780							
Qy	781	AGGACTACTGGGTAAGGACATCAACATGATGATCGGTCTCTACATCTTCCGCCAAG	840							
Db	781	AGGACTACTGGGTAAGGACATCAACATGATGATCGGTCTCTACATCTTCCGCCAAG	840							
Qy	841	TACTTGAACAGATACATGATCGGCAACTCCAGACAGATCGCTTTCACACCAACGTAAC	900							
Db	841	TACTTGAACAGATACATGATCGGCAACTCCAGACAGATCGCTTTCACACCAACGTAAC	900							
Qy	901	AACAAGCTCTCAACGAGGTTTCAAGATCATCATCAAGGATCAAGGATCAACCAAC	960							
Db	901	AACAAGCTCTCAACGAGGTTTCAAGATCATCATCAAGGATCAAGGATCAACCAAC	960							
Qy	961	GACACGAGTCAAGAGGTTGATCATCTCTTACTTCGACATGACTATCAACCAACGTC	1020							
Db	961	GACACGAGTCAAGAGGTTGATCATCTCTTACTTCGACATGACTATCAACCAACGTC	1020							
Qy	1021	TACAACCTGTTTCAAGAACGAGACCATGATACCGCCGACACCACTCCACCGGAGATC	1080							
Db	1021	TACAACCTGTTTCAAGAACGAGACCATGATACCGCCGACACCACTCCACCGGAGATC	1080							

QY 1081 TACCCATCGGTCTGCTGAGCAGACCAAGGACATCAACGAGCAACATCATCTCCAGATC 1140  
 DB 1081 TACCCATCGGTCTGCTGAGCAGACCAAGGACATCAACGAGCAACATCATCTCCAGATC 1140  
 QY 1141 CAGCCATGACCAACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGT 1200  
 DB 1141 CAGCCATGACCAACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGT 1200  
 QY 1201 GAGAACATCTCCGGTATCTGCTCCATCGGTACCTACAGATCTCGTCTGGGTGACTGG 1260  
 DB 1201 GAGAACATCTCCGGTATCTGCTCCATCGGTACCTACAGATCTCGTCTGGGTGACTGG 1260  
 QY 1261 TACAGACAACTACTTGGTTCCTCACTGTCACAGCGGTAACCTCCCTCTTGTGGAG 1320  
 DB 1261 TACAGACAACTACTTGGTTCCTCACTGTCACAGCGGTAACCTCCCTCTTGTGGAG 1320  
 QY 1321 TCCACTTCCACCACTGGGATCTGCCAGATCTCCGAGTAATAGGAATTC 1371  
 DB 1321 TCCACTTCCACCACTGGGATCTGCCAGATCTCCGAGTAATAGGAATTC 1371

RESULT 2  
 us-09-910-186a-9  
 Sequence 9, Application us/09910186A  
 GENERAL INFORMATION:  
 APPLICANT: U.S. Army Medical Research & Materiel Command  
 TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM  
 FILE REFERENCE: A33626-A 067252.0107  
 CURRENT APPLICATION NUMBER: US/09/910,186A  
 CURRENT FILING DATE: 2001-07-20  
 PRIOR APPLICATION NUMBER: PCT/US00/12890  
 PRIOR FILING DATE: 2000-05-12  
 PRIOR APPLICATION NUMBER: 09/611,419  
 PRIOR FILING DATE: 2000-07-06  
 PRIOR APPLICATION NUMBER: 60/133,865  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,866  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,867  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,868  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,869  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 60/133,873  
 PRIOR FILING DATE: 1999-05-12  
 PRIOR APPLICATION NUMBER: 08/123,975  
 PRIOR FILING DATE: 1993-09-21  
 NUMBER OF SEQ ID NOS: 34  
 SOFTWARE: Fast-Seq for Windows Version 4.0  
 SEQ ID NO 9  
 LENGTH: 1371  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 ; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum  
 ; OTHER INFORMATION: sequence  
 ; NAME/KEY: CDS  
 ; LOCATION: (10)...(1359)  
 us-09-910-186a-9

Query Match 100.0%; Score 1371; DB 34; Length 1371;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-279;  
 Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCAGATGACCACTCCCATTCATCTCTCTTACCAACCAACTCCCTGTTGAAG 60  
 DB 1 GAATTCAGATGACCACTCCCATTCATCTCTCTTACCAACCAACTCCCTGTTGAAG 60  
 QY 61 GACATCATCAGGAGTACTTCAACACATCAACGACTCCAGATCTCTCCCTGCGAGAC 120

DB 61 GACATCATCAGGAGTACTTCAACACATCAACGACTCCAGATCTCTCCCTGCGAGAC 120  
 QY 121 CGTAAGAACACCTTGTTCGACACCTCCGGTTACAAGCGCGAGGTTCTCGGAGGAGGTGAC 180  
 DB 121 CGTAAGAACACCTTGTTCGACACCTCCGGTTACAAGCGCGAGGTTCTCGGAGGAGGTGAC 180  
 QY 181 GTCCAGTGAACCAATCTTCCATTCGACTTCAAGCTGGGTTCTCCGGTGAGGACAGA 240  
 DB 181 GTCCAGTGAACCAATCTTCCATTCGACTTCAAGCTGGGTTCTCCGGTGAGGACAGA 240  
 QY 241 GGTAAAGTCTATCGTCCAGAGACGAGACATCGTCTACAACCTCCATGCTACGAGTCCCTTC 300  
 DB 241 GGTAAAGTCTATCGTCCAGAGACGAGACATCGTCTACAACCTCCATGCTACGAGTCCCTTC 300  
 QY 301 TCCATCTCTTCTTGGATCAGATCAACAAGTGGGTTCTCCAACCTGCCAGGTTACACATC 360  
 DB 301 TCCATCTCTTCTTGGATCAGATCAACAAGTGGGTTCTCCAACCTGCCAGGTTACACATC 360  
 QY 361 ATCGACTCCGTCGAAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCCCTGGTC 420  
 DB 361 ATCGACTCCGTCGAAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCCCTGGTC 420  
 QY 421 TTCACCTGAAGCAGACGAGGACTCCGAGAGTCCATCAACTTCTCTTACGACATCTCC 480  
 DB 421 TTCACCTGAAGCAGACGAGGACTCCGAGAGTCCATCAACTTCTCTTACGACATCTCC 480  
 QY 481 AACAACTCTCTGTTTACAACTGTTTCTTCTGTCACCGTCAACCAACAACTATGATGGT 540  
 DB 481 AACAACTCTCTGTTTACAACTGTTTCTTCTGTCACCGTCAACCAACAACTATGATGGT 540  
 QY 541 AACATGAAGATCTACATCAACGGTAACTGATCGACACCATCAAGGTCAAGGATGATGACC 600  
 DB 541 AACATGAAGATCTACATCAACGGTAACTGATCGACACCATCAAGGTCAAGGATGATGACC 600  
 QY 601 GGTATCAACTTCTCCAGACCATCACTTCGAGATCAACAGATCCCAACACCGGTTCTG 660  
 DB 601 GGTATCAACTTCTCCAGACCATCACTTCGAGATCAACAGATCCCAACACCGGTTCTG 660  
 QY 661 ATCACTCCGACTCCGACAACTCAACATGTGGATCCGTGACTTCTACATCTTCGCCAAG 720  
 DB 661 ATCACTCCGACTCCGACAACTCAACATGTGGATCCGTGACTTCTACATCTTCGCCAAG 720  
 QY 721 GAGTTGACGGTAAAGACATCAACATCTCTGTTCAACTCTTGCAGTACACCAACCTGCTC 780  
 DB 721 GAGTTGACGGTAAAGACATCAACATCTCTGTTCAACTCTTGCAGTACACCAACCTGCTC 780  
 QY 781 AAGGACTACTTGGGTAAAGACATCAACATCTCTGTTCAACTCTTGCAGTACACCAACCTGCTC 840  
 DB 781 AAGGACTACTTGGGTAAAGACATCAACATCTCTGTTCAACTCTTGCAGTACACCAACCTGCTC 840  
 QY 841 TACTGAACAGATACATGTAGCCACTTCCAGACAGATCGTCTTCAACACCAAGAGTAAAC 900  
 DB 841 TACTGAACAGATACATGTAGCCACTTCCAGACAGATCGTCTTCAACACCAAGAGTAAAC 900  
 QY 901 AACACGACTTCAACGAGGTTTCAAGATCATCATCAAGGATATCAGAGGTAAACACCAAC 960  
 DB 901 AACACGACTTCAACGAGGTTTCAAGATCATCATCAAGGATATCAGAGGTAAACACCAAC 960  
 QY 961 GACACAGAGTCAAGAGTGTGATCTTCTGATCTTCAAGTACTTCAACACCAAGGAGTCC 1020  
 DB 961 GACACAGAGTCAAGAGTGTGATCTTCTGATCTTCAAGTACTTCAACACCAAGGAGTCC 1020  
 QY 1021 TACAACTCTTATGAGACGAGACCATCTGTCGCGGCAACCACTCCACCGAGGACATC 1080  
 DB 1021 TACAACTCTTATGAGACGAGACCATCTGTCGCGGCAACCACTCCACCGAGGACATC 1080  
 QY 1081 TACGCCATCGGTCTGCTGAGCAGACCAAGGACATCAACGACAACTATCTTCCAGATC 1140  
 DB 1081 TACGCCATCGGTCTGCTGAGCAGACCAAGGACATCAACGACAACTATCTTCCAGATC 1140  
 QY 1141 CAGCCATGACCACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGT 1200  
 DB 1141 CAGCCATGACCACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGT 1200

QY	1201	GAGAACATCTCGGTATCTGTTCCATCGGTACCTACAGATTCGCTTGGTGGTGACTGG	1260
Db	1201	GAGAACATCTCGGTATCTGTTCCATCGGTACCTACAGATTCGCTTGGTGGTGACTGG	1260
QY	1261	TACAGACAACTACTTGGTTCCTCAACTGTCAAGCAGGTAACCTACGCCCTCTTCTGTGGAG	1320
Db	1261	TACAGACAACTACTTGGTTCCTCAACTGTCAAGCAGGTAACCTACGCCCTCTTCTGTGGAG	1320
QY	1321	TCCACTTCCAGCCACTGGGGATTGCTCCGAGTCTCCGAGTAATAGGAATTC	1371
Db	1321	TCCACTTCCAGCCACTGGGGATTGCTCCGAGTCTCCGAGTAATAGGAATTC	1371

### RESULT 3

US-09-350-756-3  
Sequence 3, Application US/09350756  
GENERAL INFORMATION:  
APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
APPLICANT: John S. Lee  
APPLICANT: Peter Pushko  
APPLICANT: Michael D. Parker  
APPLICANT: Jonathan F. Smith  
APPLICANT: Mark T. Dertzbaugh  
APPLICANT: Leonard Smith  
TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
FILE REFERENCE: 003/124/SAP RIID 98-21  
CURRENT APPLICATION NUMBER: US/09/350,756  
CURRENT FILING DATE: 1999-07-09  
EARLIER APPLICATION NUMBER: US 60/092,416  
EARLIER FILING DATE: 1998-07-10  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: Apple Macintosh Microsoft Word 6.0  
SEQ ID NO 3  
LENGTH: 1371  
TYPE: DNA  
ORGANISM: Artificial  
FEATURE:  
OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"  
US-09-350-756-3

Query Match	99.8%	Score 1366.2;	DB 17;	Length 1371;
Best Local Similarity	99.8%;	Pred. No. 3.9e-278;		
Matches 1368;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;
1	GAATTACAGGATGACCATCCGATTCACACATCTTCTCCTACACCAACAACCTCCCTGTTGAAG	60		
1	GAATTACAGATCACCATCCGATTCACACATCTTCTCCTACACCAACAACCTCCCTGTTGAAG	60		
61	GACATCATCAACGAGTACTTCAACAACATCAACGACTCCAGACTCCTGTCCTCGCAAGAC	120		
61	GACATCATCAACGAGTACTTCAACAACATCAACGACTCCAGACTCCTGTCCTCGCAAGAC	120		
121	CGTAAGACACTTGTGTCGACACTCCGGTTACAGCGCCGAGGTTCTCCGAGGAGGGTGAC	180		
121	CGTAAGAACACTTGTGTCGACACTCCGGTTACAGCGCCGAGGTTCTCCGAGGAGGGTGAC	180		
181	GTCCAGCTGAACCCAACTTCCCATTCGACTTCAAGCTGGGTTCTCCGGTCGAGGACAGA	240		
181	GTCCAGCTGAACCCAACTTCCCATTCGACTTCAAGCTGGGTTCTCCGGTCGAGGACAGA	240		
241	GGTAGGTCATCGTCACCGAGAGAGACATCGTTACAACCTCCATGTACAGATCCCTTC	300		
241	GGTAGGTCATCGTCACCGAGAGAGACATCGTTACAACCTCCATGTACAGATCCCTTC	300		
301	TCCATCTCCTTCTGATCAGAATCAACAAGTGGGTTCTCCAACCTTCCGAGGTTACACCATC	360		
301	TCCATCTCCTTCTGATCAGAATCAACAAGTGGGTTCTCCAACCTTCCGAGGTTACACCATC	360		
361	ATCGACTCCGTCAGAACAACTCCGGTTGGTCCATCCGCTATCATCTCCAACCTCCTCGTC	420		
361	ATCGACTCCGTCAGAACAACTCCGGTTGGTCCATCCGCTATCATCTCCAACCTCCTCGTC	420		

RESULT 4  
PCT-US97-15394-59  
; Sequence 59, Application PC/TUS9715394  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; TITLE OF INVENTION: Botulinum Neurotoxin

RESISTANCE

```

; PC/0597-15394-59
; Sequence 59, Application PC/TUS9715394
;
; GENERAL INFORMATION:
;
; APPLICANT: Williams, James A.
; APPLICANT: Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; TITLE OF INVENTION: Botulinum Neurotoxin
;

```



NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/15394  
FILING DATE:

## CLASSIFICATION:

## ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPD-02304  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338

## INFORMATION FOR SEQ ID NO: 59:

## SEQUENCE CHARACTERISTICS:

LENGTH: 3876 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

## FEATURE:

NAME/KEY: CDS

LOCATION: 1..3873

BASES: 15394-59

Query Match 48.5%; Score 665.2; DB 1; Length 3876;  
Best Local Similarity 68.3%; Pred. No. 4.7e-130;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

13 ACATCCCATTCACATCTCTCCCTACACACACTCCCTGTTCAAGGACATCATCAAC 72  
2527 ACAATACCCCTTTAAATATTTTTCATATAGTAAATATTTTATTAAGAGATATAAT 2586

73 GAGTACTTCAACAAACATCAACGACTCCAGATCCCTGTCGAGAACCGTAAGAACACC 132  
2587 GAATATTCATTAATATTAATGATTCATAAAATTTTGAGCTACAAACAGAAAAAT 2646

133 TTGGTCGACACCTCCGGTTACAAACGCGGAGGTCTCCGAGGAGGTGAGCTCCAGCTGAC 192  
2647 TTAGTGATACATCAGGATATATATGAAAGTGAAGAGCGGATGTTTCAGCTTAAT 2706

193 CCAATCTTCCATTCGACTTCAAGCTGGGTTCTCCGGTGGAGCAGAGGTAAAGTCAATC 252  
2707 CCAATATTCATTTGACCTTAATTTAGTTAGTTTCAGGGAGGATAGAGTTAATATA 2766

253 GTACCCAGAACAGAACATCGTTACAACTCCATGTCAGAGTCTTCTCCATCTCCTTC 312  
2767 GTACCCAGAACATCGTTACAACTCCATGTCAGAGTCTTCTCCATCTCCTTC 2826

313 TGATGATACATCAACAGTGGGTTCCCACTCCAGGTTACACCATCATCGACTCCGTC 372  
2827 TGATGATACATCAACAGTGGGTTCCCACTCCAGGTTACACCATCATCGACTCCGTC 2886

373 AAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCTCGGTCTTCCACCTGAAG 432  
2887 AAAAAAATCACTGAGTATAGGTATATTAATTTAGTAAATTTTAGTATTTACTTTAAA 2946

433 CAGAACGAGACTCCGAGCAGTCCCACTTCTCTACGACATCTCCAAACAGCTCCT 492  
2947 CAAATGAAGATAGTGAACAAAGATATAATTTAGTTATGATATCAATATCTCCT 3006

493 GGTACAAAGTGGTCTCTCGTCCAGCTCCAGCAACACATGATGGGTAAACATGAAGATC 552

3007 GGATACAATAAATGGTTTTTTTGTAACTGTTTAACTAACAATATGATGGAAATATGAAGATT 3066  
553 TACATCAACGGTTAAGCTGATCGACACCATCAAGTCAAGGAGTTGACCGGTATCAACTTC 612  
3067 TATATAAATGGAAATTAATAGTACTATAAAGTTAAAGAACTAACTGGAATATATTT 3126  
613 TCCAAGACATACCTTCGAGATCAACAAAGATCCAGACACCGGTCTGATCACTCCGAC 672  
3127 AGCAAACTATAAATTTGAATTAATAAATCCAGATACCCGGTTTGAATTTACTTCAGAT 3186  
673 TCCGACACATCAACATGTTGATCGTCTGACTTCTACATCTTCGCAAGGAGTTGGACGGT 732  
3187 TCTGATACATCAATATGTTGATAGAGATTTTATATTTTGTAAAGATATGATGGT 3246  
733 AAGGACATCAACATCTCTCAACTCTTCGAGTACCAACGCTCGTCAAGGAGTACTGG 792  
3247 AAAGATATAATATATTTAATAGCTTGAATATACTATGTTGTAAGATATTTGG 3306  
793 GGTAAAGCTGAGTACAAAGGAGTACTACATGTTCAACATCGACTTCTGAACAGA 852  
3307 GGAATGATTTAAGATATAAAGATATTTATATGTTTAAATATAGATTTTAAATAGA 3366  
853 TACATGTAAGCAACTCCAGACAGATCGTCTTCAACACGAGCTAACAACAGACTTC 912  
3367 TATGTAAGCAACTCCAGACAGATTTCTTTTAAACAGTAAATATAATGACTTC 3426  
913 AAGGAGGTACAAAGATATCAAGCTATCAGAGTATCAGAGTACACCAACGACAGATC 972  
3427 AATGAAGGATATAAATTTATATAAAGATCAGAGGAAATACAAATGATAGAGTA 3486  
973 AGAGGTGTGACATCTCTGACTTCGACATGACTATCAACAACAGGCTTCAACCTGTC 1032  
3487 CGAGGAGGAGATATTTATATTTTGCATGACATTAATAACAAGCATATAATTTGTT 3546  
1033 ATGAAGACGAGACATGTTACCCGACAAACACTCCACGAGGACATCTAGCCATCGGT 1092  
3547 ATGAAGATGAACATGTTATGATGATATCAATGATGATATATGATGATGATAGGT 3606  
1093 CTGCGTGACGACCAAGGACATCAACGACATCAATCTTCCAGATCCAGCCATGAAC 1152  
3607 TTAAGACACAAACAGGATATAATGATATATATTTCAATACCAACCAATGAAT 3666  
1153 AACATTTACTATAGCTTCCAGATCTTCAAGTCCCACTTCAAGGTTGAGAACATCTCC 1212  
3667 AATACATTTATATGACATCTCAATATTTAATCAAAATTTAATGAGAAATATTTCT 3726  
1213 GGTATCTGTTCCATCGTACAGATTTCCGTTCTGGTGGTGGTGGTGGTGGTGGTGGT 1272  
3727 GGTATATTTCAATAGTACTTATGTTTACCTTGGGTTGGGTTGGGTTGGGTTGGGTT 3786  
1273 TACTTGGTTCCAACTGTCAGAGGATCACTACGCTCTCTGCTGGAGTCCACTTCCAC 1332  
3787 TATTTGGTCTACTGTGAAGCAAGAAATTTGCTTATTATTAGTAATCAACATCAACT 3846  
1333 CACTGGGATTCGTCACCTCTCCGAGTAA 1362  
3847 CATTTGGGTTTGTACCTGTAAAGTGAATTA 3876

## RESULT 5

US-08-704-159-59

Sequence 59, Application US/08704159

GENERAL INFORMATION:

APPLICANT: Williams, James A.

APPLICANT: Thallay, Bruce S.

TITLE OF INVENTION: Multivalent Vaccine For Clostridium

TITLE OF INVENTION: Botulinum Neurotoxin

NUMBER OF SEQUENCES: 82

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/704,159  
FILING DATE:  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40, 027  
REFERENCE/DOCKET NUMBER: OPD-02304  
TELEPHONE: (415) 703-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3876 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3873  
8-704-159-59

Copied from 10910186 on 05-05-2004

Query Match 48.5%; Score 565.2; DB 11; Length 3876;  
Best Local Similarity 68.3%; Pred. No. 4.7e-130;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

13 ACATCCCATCAACATCTTCCTACACCAACAACTCCCTGTTGAAGGACATCATCAAC 72  
2527 ACATACCCCTTTAATAATTTTTCATCTACTAATAATCTTTTATAAAGATATAAT 2586  
73 GAGTACTTCAACACATCAACGACTCCAGATCCCTGCTGCGAGACCGTAAGAACCC 132  
2587 GAATATTTCAATAATTAATGATTCAAAATTTTGAGCCCTACAAAACAGAAAATACT 2646  
133 TTGCTCGACACCTCCGTTTACAAACCCGAGGTCTCCGAGGAGGTGACGTCCAGCTCAAC 192  
2647 TTAGTGATACATCAGGATATATTCGAGAGTGTGCTGAGAGCGGATGTCAGCTTAAT 2706  
193 CCAATCTCCCATTCGACTTCAAGTCGGTTCCTCCGCTGAGGACAGAGGTAAGGTCAATC 252  
2707 CCAATATTTCCATTTGACTTTAAATTTAGTAGTTCCAGGGGAGGATAGAGGTAAGTTATA 2756  
253 GTACCCAGACAGACATCTGCTACAACTCCAGTACGAGTCCCTTCCTCATCTCCTTC 312  
2767 GTACCCAGAAATGAATAATTTGTATATATTTCTATGATGAAGTTTATGAGTTAGTTT 2826  
313 TGCATCAGATCAACAAAGTGGTCTCCAACTTGCAGGTTTACACATCATCTCCGCTC 372  
2827 TGCATTAGTAATAAATAAGTGGTAACTAATTTACCTGGATATATATAAATGATAGTTT 2886  
373 AAGAACATCTCCGTTGGTCCATCGGTATCTCCCACTCCCTGGTCTTCCACCTGAG 432  
2887 AAAATATCTCAGTTGGAGTAGTAGTATTTAGTAAATTTTATGATTTTACTTTAAA 2946  
433 CAGAACGAGGACTCCGAGAGTCCCACTCTCTCTACGACATCTCCAAACAGGCTCCT 492  
2947 CAAATGAAGATAGTGAACAAAGTATAAATTTAGTTATGATATATCAAAATATGCTCCT 3006  
493 GGTACAAAGTGGTCTCTGCTACCGCTCCACACACATGATGGGTAACATGAAGATC 552  
3007 GGATCAATAAATGGTTTTTGTCTTAACTGTTACTAACAATATGATGGGAATATGAAGAT 3066  
553 TACATCAACGGTAAAGCTGATCGACACCAACATCAAGGTCACCGGATATCAACTTC 612

Db 3067 TATATAAATGGAATAATTAAGATAGTATATAAAGATTAAAGAACTAACTGGAATTAATTTT 3126  
Qy 613 TCCAAGACCATCACCTTCGAGATCAACAGATCCCGACACCGCTCTGATCACCTCCGAC 672  
Db 3127 AGCAAACTATAACATTTGAAATAAATAAATTCAGATACCGGTTTGATTTACTTCAGAT 3186  
Qy 673 TCCGACACATCAACATGTGTGATCCGTTGACTTCTACATCTTCCCAAGGAGTGGACGGT 732  
Db 3187 TCTGATACATCAATAATGTGATAAGAGATTTTATATATTTTCTAAAGAAATAGATGCT 3246  
Qy 733 AAGGACATCAACATCCCTGTTCAACTCTTCAGTACCAACGCTGCTCAAGGACTACTGG 792  
Db 3247 AAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3306  
Qy 793 GGTACAGGCTGAGATACACAGAGGATAGTACATGTGTCAACATCGACTACTTGAACAGA 852  
Db 3307 GGAATGATTTAAGATATAATAAAGAAATTAATTAATTAATTAATTAATTAATTAATTA 3366  
Qy 853 TACATGTAGCCCACTCCAGACAGATGCTTCAACACCAAGGATTAACACACGACTTTC 912  
Db 3367 TATATGTATGCAACTCAGACAAATTTGTTTAAATACAGTAAATTAATTAATTAATTAAT 3426  
Qy 913 AACGAGGTTTACAAGATCATCATCAAGCGTATCAGAGTAAACACCAACGACCCAGAGTC 972  
Db 3427 AATGAGGATATAAATTAATAAAGAAATCAGAGGAAATACAAATGATAGTAGTA 3486  
Qy 973 AGAGTGTGATCCCTGCTGATCGACATGACTATCAACACAGGCTTACACCTGTTTC 1032  
Db 3487 CGAGGAGGATATTTTATATTTTATGATGACAATTAATAACAAAGCATATAATTTGTTT 3546  
Qy 1033 ATGAAGAAGGAGACCATGTACGCGACAAACACTCCACGAGGACATCTACGCCATCGST 1092  
Db 3547 ATGAGATGAACTATGTATGCAAGTAACTAGTACTGAAGATATATATGCTATAGST 3606  
Qy 1093 CTGCTGAGGACCAAGGATCAACGACAAATCAATCTTCGAGTCCGAGCCAAATGAAC 1152  
Db 3607 TTAGAGAACAAACAAAGGATATAAATGATTAATTAATTAATTAATTAATTAATTAATTA 3666  
Qy 1153 ACACCTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGCTGAGAACATCTCC 1212  
Db 3667 AATCTTATTAATGACATCTCAATATTTAAATCAAAATTTTAAATGAGAAATATTTCT 3726  
Qy 1213 GGTATCTTCCATCGGTACCTACAGATTCCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1272  
Db 3727 GGAATATGTTCAATAGTACTATCGTTTTAGACTTGGAGGTGATTTGGTATAGACACAAT 3786  
Qy 1273 TACTGGTTCCAACTGCAAGCGGTAACTACGCTCTTGTGGAGTCCACTTCCACTCC 1332  
Db 3787 TATTGGTCCCTACTGTGAGCGAGGAAATTTATGCTTCATTTATGATTAACATCAACT 3846  
Qy 1333 CACTGGGATTCGTCCTCCAGTCTCCGAGTAA 1362  
Db 3847 CATTGGGTTTGTACCTGTAAGTGAATAA 3876

RESULT 6  
US-10-205-516-5  
; Sequence 5, Application US/10205516  
; GENERAL INFORMATION:  
; APPLICANT: Zhong, Jun  
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum  
; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique  
; FILE REFERENCE: JZBtX1  
; CURRENT APPLICATION NUMBER: US/10/205,516  
; CURRENT FILING DATE: 2002-07-25  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 3876  
; TYPE: DNA  
; ORGANISM: Clostridium botulinum  
; FEATURE:  
; NAME/KEY: CDS

LOCATION: (1), (3876)  
 PUBLICATION INFORMATION:  
 AUTHORS: Hauser, D. F.  
 TITLE: Organization of the botulinum neurotoxin C1 gene and  
 TITLE: its associated non-toxic protein genes in Clostridium  
 TITLE: botulinum C 468  
 JOURNAL: Mol. Gen. Genet.  
 VOLUME: 243  
 ISSUE: 6  
 PAGES: 631-640  
 DATABASE ACCESSION NUMBER: GeneBank / X72793  
 DATABASE ENTRY DATE: 1993-05-03  
 RELEVANT RESIDUES: 1 TO 3876  
 US-10-205-516-5

Query Match 48.5%; Score 665.2; DB 42; Length 3876;  
 Best Local Similarity 68.3%; Pred. No. 4.7e-130;  
 Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;  
 13 ACCATCCCATCAACATCTCTCCACACCAACAACTCCCTGTTGAAGCAGATCATCAAC 72  
 2527 ACAATACCCCTTAATATTTTTCATATACTAATAATCTTTTAAAGATATAATTAAT 2586  
 73 GACTACTCAACACATCAACAGCCTCCAGATCCCTGCTCCGACAGACCCGTAAGACACC 132  
 2587 GAATATTTCAATTAATTAATGATTCAAAATTTTGAAGCTCAACAAAGAAATACT 2646  
 133 TTGGTGCACACCTCCCGGTTACACGCGAGGCTCCGAGGAGGTTGAGCTCCAGCTGAAC 192  
 2647 TTAGTGATACATCAGGATATAATGCAGAGTGAGTGAAGAGGGGATTCAGCTTAAT 2706  
 193 CCAATCTCCCATTCACACTCAAGCTGGGTTCTCCGCTGAGGACAGAGGTAAGGTCATC 252  
 2707 CCAATTTTCCATTTGACTTTAAATAGTAGTTTCAGGGAGGATAGAGGTAAGTTATA 2766  
 253 GTACCCAGACAGACATCGTCTACAACTCCATGCTAGGAGTCTCTCCATCTCCCTTC 312  
 2767 GTAACCAATGAAATATTTGATATAATCTATGTAAGATTTTAGCATTAGTTT 2826  
 313 TGGATCAGATCAACAAAGTGGTCTCCAACTTGCAGGTTACACCATCATCGCTCCGTC 372  
 2827 TGGATTAGATAAATAAATGGGTAAGTAAATTTACCTGGATATATAATTCATAGTCTT 2886  
 373 AAGAAACATCCCGTGGTCCATCGGTATCATCTCCAACTTCTGCTCTCCACCTGAAG 432  
 2887 AAAAAATCACTGGTGGATAGTATGTTATTTAGTAAATTTTGTAGTTTACITTAATA 2946  
 433 CAGACAGAGACTCCGAGCAGTCCATCACTCTCCAGCAGATCTCCAAACAGCTCCCT 492  
 2947 CAAATGAAGATAGTGAACAAAGTAAATTTTGTATGATATATCAATCAATATGCTCT 3006  
 493 GGTACAAACAGTGGTCTTCGTCACCGTCACCAACATGATGGGTAACATGAAGATC 552  
 3007 GGATACAAATGAATGGTTTGTACTGTATACATATGATGGGAAATATGAAGAT 3066  
 553 TACATCAACGGTAAAGTGTGATCCAGACCATCAAGGTCAGGAGTTCACCGATCAACTTC 612  
 3067 TATATAAATGAAATTAATAGATATATAAAGTAAAGAACTAACTGGAATTAATTTT 3126  
 613 TCAACAGCATCACTCCAGATCAACAGATCCAGACACCGCTCATCACTCCAC 672  
 3127 AGCAAACTATACATTTGAAATTAATAAATTCAGATACCGGTTTGTATCTTCAGAT 3186  
 673 TCGGACAAATCAACATGGTGGATCCGCTGACTTCTACATCTTCGCAAGAGTTCGACGGT 732  
 3187 TCTGATAACATCAATATGTTGATAGAGATTTTATATATTTTGTAAAGATTTAGATGGT 3246  
 733 AAGGACATCAACATCTGTTTCACTCTCCAGTACACCAACCTGCTCAAGGACTACTGG 792  
 3247 AAGATATTAATATATTTTAACTGCTGCAATATCTAAATGTTGTAAAGATTTATGG 3306  
 793 GGTAACGACCTGAGATACAAACAGAGGAGTACTACATGGTCAACATCGACTACTTGAACAGA 852

Db 3307 GGAAATGATTTAAGATATAATAAGATATATATATGTTAATATAGATTAATTAATAGA 3366  
 QY 853 TACATGTACGCCCAACTCCAGACAGATCGTCTTCAACACCCAGCAGTACACACAGCATTC 912  
 Db 3367 TATATGTATCGCAACTCCAGCAAAATTTTAAATACACGTGAAATTAATTAATGACTTC 3426  
 QY 913 AAGCGGGTTACAAGATCATCATCAAGCGTATCAGAGGTAAACCAACGACACACAGATC 972  
 Db 3427 AATGAAGGATATAAATATATAAAGAGATCAGAGGAATACAATGATCTAGAGTA 3486  
 QY 973 AGAGTGGTGACATCTCTGACTTCGACATGACTATCAACACAGGCTTACACCTGTTTC 1032  
 Db 3487 CGAGGAGGAGATATTTTATATTTTATGATGACATTAATAAAGCAATATAATTTGTTT 3546  
 QY 1033 ATGAAGACGAGACCATGTACGCCGACACCACTCCACCGAGGACATCTACGCCATCGGT 1092  
 Db 3547 ATGAAGATTAAGATATGATGTCAGATATCATATGATGATGATATATATGCTATAGGT 3606  
 QY 1093 CTGCTGTAGCAGCAAGGACATCAACGACACATCATCTTCCAGATCCAGCCCAATGAAC 1152  
 Db 3607 TTAAGAGAAACAACAAAGGATATAAATGATATATATTTTCAATACAAACCAATGAAT 3666  
 QY 1153 AACACATCTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGAACATCTCC 1212  
 Db 3667 AATCTATATATACGATCTCAATATTTAAATCAAAATTTTAIGGAGAAATTTCT 3726  
 QY 1213 GGTATCTGTCCATCGGTACCTACGATTCGCTCTGGGTGGTGTGCTGTAGCAGACACAA 1272  
 Db 3727 GGAATATGTTCAATAGTACTTATGCTTTAGACTTGGAGGTGATTTGATATAGACAAAT 3786  
 QY 1273 TACTTGTCTCCAACTGTCAAGCAGGTAACGCTCTCTGCTGGAGTCCACTCCAC 1332  
 Db 3787 TATTGCTGCTTACTGTGAAGCAAGAAATTAATGCTTATTATAGAAATCAACATCACT 3846  
 QY 1333 CACTGGGATTCGTCACGATCTCCGAGTAA 1362  
 Db 3847 CATTTGGGTTTGTACTCTGTAAGTGAATAA 3876

RESULT 7  
 US-10-271-012-59  
 ; Sequence 59, Application US/10271012  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Williams, James A.  
 ; ; Thalley, Bruce S.  
 ; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
 ; ; Botulinum Neurotoxin  
 ; NUMBER OF SEQUENCES: 82  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Medlen & Carroll  
 ; STREET: 220 Montgomery Street, Suite 2200  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States of America  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/271,012  
 ; FILING DATE: 15-Oct-2002  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/704,159  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ingolia, Diane E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: OPD-02304  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 59:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3876 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..3873  
 SEQUENCE DESCRIPTION: SEQ ID NO: 59:  
 US-10-271-012-59

Query Match 48.58; Score 665.2; DB 43; Length 3876;  
 Best Local Similarity 68.34; Pred. No. 4.7e-130;  
 Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

13 ACCATCCCATTCACATCTTCTCCTACCAACAACCTCCCTGTTGAGGACATCATCAAC 72  
 2527 ACAATACCCITTAATATTTTTCATATCTACTANPAATTTCTTTNTAAAGATATAAT 2586

73 GAGTACTTCAACACATCAACGATCCAGATCTGTCCTCGAGACCGTAAAGACACC 132  
 2587 GAATATTTCAATATATATGATGATCAAAAATTTTGAGCCTACAAACAGAAAAATACT 2646

133 TTGGTCGACACTCCGGTTACAACGCCGAGGTCTCCGAGGAGGTGAGCTCCAGCTGAAC 192  
 2647 TTAGTGATACATCAGGATATATGAGAGAGTGTGAGAGAGCGATGTCAGCTTAAT 2706

193 CCAATCTTCCATTCGACTTCAAGCTGGTTCCTCCGGTGAGGACAGAGTAAAGTCATC 252  
 2707 CCAATATTTCCATTTGACTTTAAATAGTAGTGTTCAGGGAGGAGTAAAGTTATA 2766

253 GTACCCAGAGAGAGACATGCTCAACTCCATGACGATGACGCTTCTCCATCTCCTTC 312  
 2767 GTACCCAGAGATATATGATATATATCTATGATGATGATGATGATGATGATGATGAT 2826

313 TGGATCAGAAATCAAAAGTGGGTCCCAACTGCCAGTTTACACCATCATCGACTCCGTC 372  
 2827 TGGATTAATAATAATAATAGTAAAGTAAATTTACCTGGATATATATATGATGATGAT 2886

373 AAGAACACTCCGGTGTGCTCATCGATCATCTCCAACTTCTGGTCTTCCCTCAAG 432  
 2887 AAAATTAATCAAGTGTGAGTATGATATATGATGATGATGATGATGATGATGATGAT 2946

433 CAGAACGAGGACTCCGAGCAGTCCATCACTTCCTTACGACATCTCCAAACAGCTCCT 492  
 2947 CAAATGAAGATAGTGAACAAAGTATAAATTTTATGATGATGATGATGATGATGATGAT 3006

493 GGTACACAAAGTGGTCTTCTGTCACGTCACCAACACATGATGGTACATGAAGATC 552  
 3007 GGATACAAATAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3066

553 TACATCAACGCTAGCTGATCGACACCATCAAGTCAAGGAGTGTGACCGGTATCAACTTC 612  
 3067 TATATAATGGAATAATTAATAGATATATAAAGTTAAAGAACTAACTGGAATTAATTTT 3126

613 TCCAGACCATCACTTCCGATCAACAGATCCGACACCGGTCTGATCACCCTCCGAC 672  
 3127 AGCAAACTATAAATTTGAATAATAAATAATCCAGATACCGGTTTGTATCTTCAGAT 3186

673 TCCGACACATCAACATGATGGATCCGCTGACTTCTACATCTTCGCAAGGAGTGTGGACGGT 732  
 3187 TCTGATACATCAATATGATGGATGAAGATTTTATATATTTGCTAAGATATGATGGT 3246

733 AAGGACATCAACATCTCTTCACTCTTCCAGTACACCAAGCTCGTCAAGGACTACTGG 792  
 3247 AAAGATATTAATATTTTAAATAGTGTGCAATATACTAATGTTGTTAAAGATTTATGG 3306

793 GGTACGACCTGAGATACAAAGAGGTACTACATGTTGTCACATCGGACTACTTGAACAGA 852  
 3307 GGAATGATTTAAGATATATAAAGATATATATGATGATGATGATGATGATGATGATGAT 926

QY 853 TACATGTACGCCCAACTCCACAGACAGATCGCTTCCAAACACGACGAGTACAAACAGCACTTC 912  
 DB 3367 TATATGTATGCGAACTCAGCAAAATTTGTTTTTAATACACGCTAGAAATTAATATGACTTC 3426

QY 913 AACGAGGTTTACAGATCATCATCAAGCGGTATCAGAGGTAAACACCAACGACACAGAGTC 972  
 DB 3427 AATGAAGATATAAATTAATAAAGAAATCAGAGAAATACAAATGATAGAGTA 3486

QY 973 AGAGTGGTGACATCTCTGTACTTCGACATGATCATCAACAAAGGCCCTACAACTGTTC 1032  
 DB 3487 CGAGGAGGATATTTTATATTTTATGATGACAAATTAATAACAAAGCATATAATTTGTT 3546

QY 1033 ATGAAGACGAGACCATGTACGCGCAACCACTCCACGAGGACATCTACGCCATCGGT 1092  
 DB 3547 ATGAAGATGAAGTATGATGACATATATCATATGATGATGATGATGATGATGATGAT 3606

QY 1093 CTGCGTGAGCAGACCAAGGACATCAACGACACATCATCTTCCAGATCCAGCAATGAAC 1152  
 DB 3607 TTAAGAGAACAAACAAAGGATATAAATGATATATATATTTTCAAAATACACCAATGAAT 3666

QY 1153 AACACTTACTACTACGCTCCGAGATCTTCAAGTCCAACTTCAACGCTGAGAACATCTCC 1212  
 DB 3667 AATCTTATTTATACGCTCTCAATATTTTAAATCAAAATTTTATGGAGAAATATTTCT 3726

QY 1213 GGTATCTGTTCCATCGGTACTTACAGATTCGCTCTGGGTGGTACTGTTACGACACAAAC 1272  
 DB 3727 GGAATATGTTCAATAGTACTTATCGTTTTAGACTTTGGAGGTGATGTTGATAGACAAAT 3786

QY 1273 TACTTGGTCCAACTGTCAGCAGGCTAAGTACGCTTCTGCTGGAGTCCACTTCCCTCCACC 1332  
 DB 3787 TATTGGTGGCTACTGTGAAGCAAGGAAATTTGCTTCATTTATTAGATCAACATCAACT 3846

QY 1333 CACTGGGGAATCGTCCCACTCCCGAGTAA 1362  
 DB 3847 CATGGGGTTTGTACCTGTAAAGTGAATA 3876

RESULT 8  
 US-08-954-302-2  
 ; Sequence 2, Application US/08954302  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lance Simpson, Nikita Kiyatkin,  
 ; APPLICANT: Andrew Maksymowich  
 ; TITLE OF INVENTION: Compositions and Methods for Systemic  
 ; DELIVERY OF ORAL VACCINES AND THERAPEUTIC AGENTS  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Jane Massey Licata, Esq.  
 ; STREET: 66 E. Main Street  
 ; CITY: Marlton  
 ; STATE: NJ  
 ; COUNTRY: USA  
 ; ZIP: 08053  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
 ; COMPUTER: IBM 486  
 ; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/954,302  
 ; FILING DATE: herewith  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jane Massey Licata  
 ; REGISTRATION NUMBER: 32,257  
 ; REFERENCE/DOCKET NUMBER: JEFF-0164  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (609) 779-2400  
 ; TELEFAX: (609) 810-1454

```

; INFORMATION FOR SEQ ID NO: 2:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 3950
;
; TYPE: NUCLEIC ACID
;
; STRANDEDNESS: SINGLE
;
; TOPOLOGY: LINEAR
;
; ANTI-SENSE: NO
;
DS-08-954-302-2

```

Query Match	48.5%;	Score 665.2;	DB 13;	Length 3950;
Best Local Similarity	58.3%;	Pred. No. 4.7e-130;		
Matches 922;	Conservative	0;	Mismatches 428;	Indels 0;
Gaps	0;			

13 ACATCCCATTCAGACATCTTCTCCTACACCAACAACTCCCTGTTGAGGACATCATCAAC 72

2558 ACAAACCCCTTAAATATTTTTCATATATACTAATCAATCTTTATTTAAAGGATATAATTAAT 3617

73 GAGTACTTCAACAACATCAACGACTCCAGACTCTCTGCTGCGAGAACCCTGATAGAACAAC 132

2618 GAATATTTCAATAATTAATGATTCAAAATTTTGACCTTACAAAACAGAAAATACT 2677

133 TTGGTGCACACCTCCCGGTTACAAACCCGAGGTCCTCCGAGGAGGTGACGTCCACGCTGAAC 192

2678 TTAGTGGATACATCAGGATATTAATCAGAAGTGAGTGAAGAAGCGGATGTTCAAGCTTAAT 2737

193 CCAATCTTCCCATTCGACATTCAGCTGGGTCCCTCCGGTGAGGACAGAGTAAGGTCATC 352

2738 CCAATATTTCCATTGACITTAATAGGTAGTTCAGGGAGGATAGAGTAAAGTATATA 2797

253 GTACCCAGAACGAGAACATCGCTTACAACTCCAGTACAGAGTCCCTCTCCATCTCCTCTC 312

2798 GTAAACCCAGAAATATGATATATAATCTATGATGAAGCTTTTTCAGCTAGTATTT 2857

313 TGGATCAGAAATCAACAAGTGGGTCCCAACTTGCCAGGTTACACCATCATCGACTCGCTC 372

2858 TGGATTAGATAAATAAATGGGTAGTAAATTTACCTGGATATACATAATGATAGTGT 2917

373 AAGAACACTCCCGTTGGTCCATCCGGTATCATCTCCAACTTCTCTGGTCTTCAACCTGAAG 432

2918 AAAAATAACTCAGGTGGAGTAGTAGTATTATTAGTAATTTTTTAGTATTACTTTAAAA 2977

433 CAGAAGCAGGACTCCGAGCAGTCCCAACTCTCTACGACATCTCCAAACAACGCTCCT 492

2978 CAATATGAAGATAGTAGAACAAAGTATAATTTAGTTATGATATATCAATATGCTCCT 3037

493 GGTTCACAAAGTGGTTCTTCGTCCACCGTCCACCAACAACATGATGGGTACATGAAGATC 552

3038 GGATACAATAAATGGTTTTTTGTACTGTTTACTAACAAATATGATGGGAAATATGAAGATT 3097

553 TACATCAACGGTAACTGATCGACACCATCAAGGTCAAGGAGTTGACGGGTATCAACTTC 612

3098 TATATAATCGAAATTAATAGATCTATAAAGTTAAAGACTAACTGGANTTAATTTTT 3157

613 TCCAAGACCATCACCTTCGAGATCAACAGATCCCGAGACACCGGTCTGATCACTCCGAC 672

3158 AGCAAAACTATAACATTTGAATATAATAAAATCCAGATACCGGTTGATTACTTTCAGAT 3217

673 TCCGACACATCAACATGTGGATCCGTGACTTCTACATCTTCGCCACAGGAGTTGCGACGT 732

3218 TCTGATAACATCAATATGTTGATAGAGATTTTTATATTTGCTAAGAATTAAGTGGT 3277

733 AAGGACATCAACATCTCTTCAACTCCTTGCAGTACACCAACGCTGTCGAAGGACTACTGG 792

3278 AAAGATATTAATATATTTAATAGCTTGCATATATACTAATGTTGTAAGAAGATTATTGG 3337

793 GGTACACACCTGAGATACAAACAGGAGTACTACATGGTTCACATCGACTACTTGAACAGA 852

3338 GGAATGATTAAAGATATAATAAAGAATATTATATGGTTAATAGATTATTTAAATAGA 3397

853 TACATGTACGCCAACTCCAGACAGATCTCTTCAACACCCAGACAGTAAACAACAGGACTTC 912

3398 TATATGTATCGGAACCTCAGCAAAATTTGTTTTTANTACAGCTAGAAAAATAATAGACTTC 3457

RESULT 9  
US-10-205-516-19  
; Sequence 19, Application US/10205516

GENERAL INFORMATION:  
 APPLICANT: Zhong, Jun  
 TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum Neurotoxins Through Recombinant DNA technique  
 FILE REFERENCE: JZbtx1  
 CURRENT APPLICATION NUMBER: US/10/205,516  
 CURRENT FILING DATE: 2002-07-25  
 NUMBER OF SEQ ID NOS: 28  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 19  
 LENGTH: 3906  
 TYPE: DNA  
 ORGANISM: Clostridium botulinum  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)..(3906)  
 FEATURE:  
 NAME/KEY: misc\_signal  
 LOCATION: (1347)..(1358)  
 OTHER INFORMATION: factor xa site  
 FEATURE:  
 NAME/KEY: misc\_signal  
 LOCATION: (3886)..(3903)  
 OTHER INFORMATION: 6-histidine tag  
 US-10-205-516-19

Query Match	48.48;	Score 563.2;	DB 42;	Length 3906;
Best Local Similarity	68.48;	Pred. No. 1.2e-129;		
Matches 917: Conservative	0;	Mismatches 423;	Indels	0; Gaps 0;

Qy	13	ACGATCCATTCACAACTCTCTCCTTACACCAACACTCCCTGCTTCAGACGACATCATCTCAAC	72
Db	2339	ACAATACCCCTTAAATATTTTCTCTACTCAATATATCTTTATTAAGAGATATATATTAAT	2598
Qy	73	GAGTACTTCACCAACTCAACGACTCCAGACTCTCGCTCGCAGACCGTAAAGAACCC	132
Db	2599	GAAATATTTCATATATATTAATGATTCATAAATTTTGAAGCTTCACAAACGAAAAAATACT	2658

QY	133	TTGGTCGACCTCGGTTACAACSCOGAGTCTCCGAGGAGGTGACGTCCAGCTGAAC	192
Db			
Db	2659	TTAGTGGATACATCAGGATATAATGCAGAACTGAGTGAAGAAGCGGATGTTCACTTAAT	2718
QY	193	CCAACTCTCCCATTCGACTTCAAGCTGGTTCCCTCCGGTGAGGACAGAGTAAGTCAATC	252
Db			
Db	2719	CCAAATATTTCCATTGACTTTAAATAGGTAGTTACGGGGAGGATAGAGTAAGTTATA	2779
QY	253	GTCAACCCAGAAATGAATAATTTGTATATAATTTCTATGTATGAAAGTTTGTAGCTTTT	2838
Db			
QY	313	TGGATCAGAACTCAACAAGTGGTCTCCAACCTTGCAGGTTACACCATCATCGACTCGCT	372
Db			
Db	2839	TGATTAGAAATAAATGGGTAGTAAATTTTACCTGGATATCTAATTTGATAGTGT	2898
QY	373	AAGAACAACCTCGGTTGGTCCATCGGTATCAATCTCCAACCTTCTCGTCTTCAACCTGAAG	432
Db			
Db	2899	AAAAATAACTCAGGTGGAGTATAGTATTAATAGTAATTTTTTAGTAATTTTACTTTAAAA	2958
QY	433	CAGACGAGGACTCCGACAGTCCATCAACTTCTCCTACGACATCTCCAACAAGCTCCT	492
Db			
Db	2959	CAAAATGAAGATAGTGACAAAGATAAATTTTAGTTATGATATATCAAAATAATGCTCCT	3018
QY	493	GGTTACACAAGTGGTTCTTCCTCACCCTGCACCAACATGATGGTGAACATGAAGATC	552
Db			
Db	3019	GGATACAATAAATGGTTTTTTTGTAACTGTTACTTAACAATATGATGGGAAATATGAAGATT	3078
QY	553	TACATCAACGGTAAAGTCTGTCGACACCATCAAGTCTCAAGNGTGCACGGTATCAACTTC	612
Db			
Db	3079	TATATAATGGAATAATATAGTACTATAAAGATTAAAGACTTAACCTGGAAATAATTTT	3138
QY	613	TCCAAGACCATCACCTTCGAGATCAACAAGATCCAGACACCGCTCTGATCACTCCGAC	672
Db			
Db	3139	AGCAAAACTATAAATTTGAAATTAATAAATTCAGATAACCGTTTTCATTACTTCAGAT	3198
QY	673	TCCGACACATCAACATGTGGTCCGTGACTTCTACATCTTCGCCAAGGAGTTGGACGGT	732
Db			
Db	3199	CTGTAAACATCAATGTGTGATAGAGATTTTATATTTTGTGAAGAATATAGATGGT	3258
QY	733	AAGGACATCAACATCCTGTTCAACTCCTTCGAGTACACCAACGTCGTCAGGACTACTGG	792
Db			
Db	3259	AAAGATATTAATATATTTTAAAGCTTGCATATCTAATGTTGTAAAGATTTATGG	3318
QY	793	GGTAACGACCTGAGATACAAACAAGGAGTACTACATGCTCACTCGACTACTTGAACAGA	852
Db			
Db	3319	GGAAATGATTTTAAGATATAATAAAGAAATATTATATGTTAATATAGATTTTAAATAGA	3378
QY	853	TACATGTACGCCAATCCAGACAGATCGTCTTCAACACCAGACGTAACAACAACGACTTC	912
Db			
Db	3379	TATATGTATGCCAATCCAGCAAAATGTTTTTAATACAGTAGAATAATAATGACTTC	3438
QY	913	AACGAGGTTACAGATCATCATCNAGGCTATCAGAGTACACCCACGACACACGAGTTC	972
Db			
Db	3439	AATGAAGGATATAAAATATTATAAAGAAATCAGAGAAATCAAAATGATCATAGATA	3498
QY	973	AGAGTGGTGACATCCTGTACTTCGACATGACTATCAACACAAAGGCTTCAACCTGTTTC	1032
Db			
Db	3499	CGAGGAGGAGATATTTTATATTTTGATATGACAAATTAATACAAAGCATATAATTTGTTT	3558
QY	1033	ATGAAGACGACACCATGTAGCCGACACACCATCTCCACCGAGGACATCTTACGCCATCGGT	1092
Db			
Db	3559	ATGAAGAATGAACATATGTATGACAGATAATCATAGTACTTGAAGATATATATGCTATAGGT	3618
QY	1093	CTCGGTGAGCACCCCAAGGACATCAACGACAAACATCATCTTCCAGATCCAGCCAAATGAAC	1152
Db			
Db	3619	TTAAGAGAACAAACAAGGATATAAATGATAAATTAATTTCAATACAAACCAATGAAT	3678
QY	1153	ACACTTACTACTACGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGACATCTCC	1212
Db			
Db	3679	AATACTTATATTAACGCATCTCAAAATATTTAAATCAAAATTTTATGAGGAAATATTTCT	3738









Db 1255 ATAATATTATTTCATCAATACAAACCAATGAAATATCTTTATTATTACGATCTCAATAT 1314  
 QY 1181 TCAAGTCCAACTTCAACGGTGAGAACATCTCCGGTATCTGTTCCATCGGTACTACAGAT 1240  
 Db 1315 TTAATCAAAATTTATGAGAGAAATATTTCGATATGTTCAATAGGTACTTATCGTT 1374  
 QY 1241 TCGGTCTGGGTGCTGCTGCTAGACAGACAACTACTCTGTTCCAACTGTCAGAGGGA 1300  
 Db 1375 TTAGACTGGAGGTGATGTTGATATAGACAAATATTATTGTTGCTACTGTGAAGCAAGAA 1434  
 QY 1301 ACTACGGCTCTTCTGCGAGTCTCCATCTCCACCCACTCGGGGATCTCGGAGT 1360  
 Db 1435 ATATGCTGCTATTTAGATCAACATCACTCACTTGGGGTTTGTACTGTGAAGTGAAT 1494  
 QY 1361 AATAG 1365  
 1495 AAAAG 1499

## RESULT 13

US-08-704-159-61  
 Sequence 61, Application US/08704159  
 GENERAL INFORMATION:  
 APPLICANT: Williams, James A.  
 APPLICANT: Thalley, Bruce S.  
 TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
 TITLE OF INVENTION: Botulinum Neurotoxin  
 NUMBER OF SEQUENCES: 82  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Medlen & Carroll  
 STREET: 220 Montgomery Street, Suite 2200  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94104

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/704,159  
 FILING DATE:

CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ingolia, Diane E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: OPD-02304

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 61:

SEQUENCE CHARACTERISTICS:

LENGTH: 1502 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 108..1493

US-08-704-159-61

Query Match 47.4%; Score 649.8; DB 11; Length 1502;  
 Best Local Similarity 68.2%; Pred. No. 7.1e-127;  
 Matches 903; Conservative 0; Mismatches 422; Indels 0; Gaps 0;

QY 41 CCAACAACTCCCTGTTGAAGGACATCATCAACGAGTCTCAACAACTCAACAGACTCCA 100

Db 175 CTAGCATGGCTTTATTAAGAGATATATTAAGAGATATTAAGAGATATTAAGAGAT 234

QY 101 AGATCCCTGCTCCGAGAACCGGTAGAACACCTTGTGTCACACCTCCGGTTACACGCG 160

Db 1315 TTAATCAAAATTTTAATGGAGAAATATTTCTGGAATATGTTCAATAGTACTTATCGTT 1374

Db 235 AAATTTTGTAGCCTTACAAAACAGAAAAATACATTTAGTGGATACATCAGGATATATGCG 294  
 QY 161 AGTCTCCGAGGAGGTGAGCTCCAGCTGAACCCCAATCTTCCATTCGACTTCAAGCTGG 220  
 Db 295 AAGTGTAGTGAAGAGCGATGTTTCACTTTAATCCCAATATTTCCATTTGACTTTAAATAG 354  
 QY 221 GTTCTCCGGTGAAGAGAGGTGAGGTGATCGTCAACCCAGAGAACGAGAACATCGCTTACA 280  
 Db 355 GTAGTTTCAAGGGAGGATAGAGGTTAAGTTATAGTAACCCAGAAATGAAAAATATTGTATATA 414  
 QY 281 ACTCCATGTACGAGTCTTCTCCATCTCTCTGGATCAGATCAGATCAGAAAGTGGGCTCCA 340  
 Db 415 ATTCTATGTATGAAGATTTTAGCATTTAGTTTGGATAGTAATAATAATGGGTAAGTA 474  
 QY 341 ACTTGCAGGTTTACACCATCATCGACTCCGCTCAAGAAACAACTCCGGTTGGTCCATCGGTA 400  
 Db 475 ATTTACCTGGATATCTATAATTTAGTTTAAATAAATCACTCAGTTGGAGTATAGTA 534  
 QY 401 TCATCTCCAACTTCTCTGCTTCCACCTGAAGCAGAACGAGGACTCCGAGGAGTCCATCA 460  
 Db 535 TTATTAGTAATTTTGTAGTATTACTTTTAAACAAAATGAAGATAGTAGAACAAAGTATAA 594  
 QY 461 ACTTCTCCTACGACATCTCCAAACACGCTCCCTGGTTTCAACAAAGTGGTTCTTCTCGTCA 520  
 Db 595 ATTTTACTGATATATCAATATGCTCTCTGGATACAAATAATGGTTTCTTGTAACTG 654  
 QY 521 TCACCAACACATGATGGGTAAATGATCTACATCAACCGGTAAAGCTGATCGACACCA 580  
 Db 655 TTACTAACATATGATGGGAAATATGAGATTTATATAATGGAATAATTAATAGTACTA 714  
 QY 581 TCAAGGTCAAGGAGTTGACGGTATCAACTTCTCAAGAACCATCACCTTCGAGATCAACA 640  
 Db 715 TAAAGTTAAGAACTAAGTGAATTAATTTAGCAAACTATACATTTGAAATAATA 774  
 QY 641 AGATCCAGACACGGTCTGATCACCTCCGACTCCGACAACTCAACATGTGGATCCCGT 700  
 Db 775 AAATTCACATACCGGTTTGTACTTCTAGATTTCTGATTAACATCAATATGTGGTAAG 834  
 QY 701 ACTTCTACATCTTCCCAAGGAGTTGGAGTAAAGACATCAACATCTGTTGACTCCT 760  
 Db 835 ATTTTATATATTTGCTAAAGAAATTAAGATGTTAAAGATATTAATATATTTAATAGCT 894  
 QY 761 TGCAGTACACCAACGCTCGTCAAGGACTACTGGGTAAAGACCTGAGATACAAACAGGAGT 820  
 Db 895 TGAATATATACTAATGTTGTAAGAGATTTTGGGAAATGATTTAAGATATAATAAAGAT 954  
 QY 821 ACTACATGTTCAACATCGACTACTTCAACAGATACATGTACGCCAACTCCAGACAGATCG 880  
 Db 955 ATTATATGTTAATATAGATTTTAAATAGATATATGTATCGGAACCTACAGCAAAATG 1014  
 QY 881 TCTTCAACACACGAGCTTCAACACAGCTTCAACAGGGTTTCAAGATCATCATCAAGC 940  
 Db 1015 TTTTAAATACACGTAGAAATAATATGACTTCAATCAAGGATATAAATAATATAATAAAA 1074  
 QY 941 GTATCAGAGGTAAACACAGGACACGAGTCAAGGTTGGTGCATCTGTTACTTCGACA 1000  
 Db 1075 GAATCAGAGGAAATACAAATGATAGTACGAGGAGGAGATATTTTATATTTTGATA 1134  
 QY 1001 TGACTATCAACAAAGGCTCAACCTGTTTCAATGAAGAACGAGACCATGTACGCCGACA 1060  
 Db 1135 TGACAATTAATACAAAGCATATAATTTGTTTATGAAGAAATGAACACTATGTATGCGATA 1194  
 QY 1061 ACACCTCCAGGAGGACATCTAGCCCATCGGTCCTGGGTGAGCAGACCAAGGACATCAACG 1120  
 Db 1195 ATCATAGTACTGAAGATATATATGCTATAGTTTAAAGAGAACAAACAAAGGATATAAATG 1254  
 QY 1121 ACAACATCATCTTCCAGATCCAGCCCAATGAACAACTTACTTACTACGCTTCCAGATCT 1180  
 Db 1255 ATAAATATATTTCAAAATACAACTCAATGAATATATTTATTTACGATCTCAATAT 1314  
 QY 1181 TCAAGTCCAACTTCAACGGGTGAGAACATCTCCGGTATCTGTTCCATCGGTACCTACAGAT 1240  
 Db 1315 TTAATCAAAATTTTAATGGAGAAATATTTCTGGAATATGTTCAATAGTACTTATCGTT 1374

[illegible]

## RESULT 14

475	ATTTACCTTGGATATACATAATCTGNTAGTGTAAATAA	ACTCAGGTGGAGTATAGGTA	534
401	TCATCTCAACTTCCTGGCTTCACCCGTGAAGACAGAGGACTCGGACGACTGCATCA	460	
535	TTAATTAGTAATTTTTTTAGTATTTACTTTAAACAAATGAAGTAGTGAACAAAGTATTA	594	
461	ACTTCTCCTACGACATCTCCAAACACGCTCCTGCTTACAACAAGTGGTCTTCCGTCCACCG	520	
595	ATTTTGTATGATATCAAAATATGCTCCTGGATACAATAAATGGTTTTTGTAACTG	654	
521	TCACCAACAACATGATGGGTAAACATGAAGATCTACATCAACGGTAAAGCTGATCGACACCA	580	
655	TTACTTAACAATATGATGGGAATATGAAGATTTTATATAATGGAAATTAATAGATACTA	714	
581	TCAAAGGTCAAGGATGACCGGTATCAACTTCTCCAAGACATCACCTTCGAGATCAACA	640	
715	TAAAGTTTAAAGAACTAACTGGAATTAATTTTACAAACTATAACATTTGAATAAATA	774	
641	AGATCCCAACACCGGCTGATACACTCCGACTCCGACACACATCAACATGTGGATCGGTG	700	
775	AAATTTCCAGATACCGGTTTGATTTACTTCAGATCTCGATAACATCAATATGTGGATAAGAG	834	
701	ACTTCTACATCTTCGCCAAGAGTTGGACGGTAAAGGACATCAACATCCTGTTCAACTCCT	760	
835	ATTTTATATATTTGCTTAAAGATTAAGTGTAAAGATATTAATATATTATTTAATAGCT	894	
761	TGCATGACACCAACGCTCGTAAGGACTACTCGGGGTACGACCTGAGATACAAACAGAGT	820	
895	TGCAATATACTAATGTTGTAAGAATTAATTCGGGAATGATTTAAGATATAATAAAGAAT	954	
821	ACTACATGTCACCATCGACTACTTGTAACACATACATGTACGGCCACTCCAGACAGATCG	880	
955	ATTATATGGTTAATATAGATTTATTTAAATAGATATGTATGCCACTCAGACAAATTC	1014	
881	TCCTCAACACGAGCGTAAACAACAGCACTTCAACGAGGGTTACAGATCATCATCAAGC	940	
1015	TTTTTAAATACGTAAGAATAATAATGACTTCAATGAAGGATATAAAATATATAAATAA	1074	
941	GTATCAGAGGTAAACCAACACACACGAGTGCAGAGTGGTGACATCCTCTACTTCGACA	1000	
1075	GAATCAGAGGAATAACAATGATCTAGATACGAGGAGGATATTTTATATTTTCATTA	1134	
1001	TGACTATCAACAACAGGCCCTACAACCTGTTTCATGAAGAACGAGACCATGTACCGGACA	1060	
1135	TGACAATTAATAACAACACATATAATTTGTTTATGAAGAATGAACACTGATGTATGCGATA	1194	

1061	QY	ACCACTCCACCGAGACATCTACGCCATCGCTCTGCTGAGCAGACCAAGACATCAACG	1120
1195	Db	ATCATAGTACTGAAGTATATATGCTATAGGTTTAAAGACAACAAGAGATATAAATG	1254
1121	QY	ACAACATCATCTTCACAGATCCAGGCCAATGAACACACTTTACTACGCTTCCACAGATCT	1180
1255	Db	ATAATATATATTTTCAATCAACAACATGAATATACTTATTTACGCACTCAAAATAT	1314
1181	QY	TCAAGTCCAACHTTCAACGGTGGAGAACATCTCCGGTATCTGTTCCATCGGTACTCAGAT	1240
1315	Db	TTAAATCAAAATTTTAAATGGAGAAAATATTTCTGGAATATGTTCAATAGGTACTATCGTT	1374



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: November 7, 2002, 12:26:05 ; Search time 34 Seconds  
(without alignments)  
4435.920 Million cell updates/sec

Title: US-09-910-186A-9  
Perfect score: 1371  
Sequence: 1 gaattccgagaccatccc.....ttcccgagtaagggaattc 1371

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 193892 seqs, 55004114 residues

Total number of hits satisfying chosen parameters: 387784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_New:\*  
1: /cgn2\_6/ptodata/1/pna/US06\_NEW\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/pna/US10\_NEW\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/pna/US11\_NEW\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/pna/US12\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	6.9	95	135259	6	US-10-240-425-1585
2	58.2	4.2	3028	5	US-09-548-938A-2
3	50.6	3.7	2835	6	US-10-092-411A-1515
4	50.6	3.7	3561	6	US-10-092-411A-1685
5	50.6	3.7	6414	6	US-10-092-411A-1626
6	50.4	3.7	15645	6	US-10-085-198-111
7	48.4	3.5	1452	6	US-10-270-223-1
8	47.6	3.5	783	6	US-10-092-411A-1670
9	46.8	3.4	8249	6	US-10-240-955-138
10	46	3.4	2456	6	US-10-267-255-83
11	46	3.4	18034	6	US-10-267-255-75
12	45.8	3.3	1092	6	US-10-270-223-5
13	45.6	3.3	1821	6	US-10-267-255-82
14	44.4	3.2	3401	6	US-10-230-437-123
15	44.4	3.2	3401	6	US-10-125-923A-411
16	43.2	3.2	8833	6	US-10-240-425-1410
17	42.8	3.1	2175	6	US-10-267-255-85
18	42.8	3.1	3878	6	US-10-264-237-907
19	42.8	3.1	4174	6	US-10-194-967-5
20	41.8	3.0	2334	6	US-10-131-813A-129
21	41.8	3.0	2334	6	US-10-131-819A-129
22	41.8	3.0	2334	6	US-10-131-823A-129
23	41.8	3.0	2334	6	US-10-131-824A-129
24	41.8	3.0	2334	6	US-10-131-826A-129
25	41.8	3.0	2334	6	US-10-131-829A-129
26	41.8	3.0	2334	6	US-10-125-926A-129

27	41.8	3.0	2334	6	US-10-127-829A-129	Sequence 129, App
28	41.8	3.0	2334	6	US-10-127-831A-129	Sequence 129, App
29	41.8	3.0	2334	6	US-10-127-835A-129	Sequence 129, App
30	41.8	3.0	2334	6	US-10-127-837A-129	Sequence 129, App
31	41.8	3.0	2334	6	US-10-127-842A-129	Sequence 129, App
32	41.8	3.0	2334	6	US-10-127-850A-129	Sequence 129, App
33	41.8	3.0	2334	6	US-10-127-901A-129	Sequence 129, App
34	41.8	3.0	2334	6	US-10-128-889A-129	Sequence 129, App
35	41.8	3.0	2334	6	US-10-131-830A-129	Sequence 129, App
36	41.8	3.0	2334	6	US-10-131-833A-129	Sequence 129, App
37	41.8	3.0	2334	6	US-10-131-837A-129	Sequence 129, App
38	41.8	3.0	2334	6	US-10-125-930A-129	Sequence 129, App
39	41.8	3.0	2334	6	US-10-127-825A-129	Sequence 129, App
40	41.8	3.0	2334	6	US-10-127-838B-129	Sequence 129, App
41	41.8	3.0	2334	6	US-10-127-843A-129	Sequence 129, App
42	41.8	3.0	2334	6	US-10-127-849A-129	Sequence 129, App
43	41.8	3.0	2334	6	US-10-128-684A-129	Sequence 129, App
44	41.8	3.0	2334	6	US-10-128-685A-129	Sequence 129, App
45	41.8	3.0	2334	6	US-10-128-686A-129	Sequence 129, App

## ALIGNMENTS

RESULT 1  
US-10-240-425-1585  
; Sequence 1585, Application US/10240425  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Amanda  
; APPLICANT: Boland, Joseph F.  
; APPLICANT: Lord, Reginald V.  
; APPLICANT: Alvarez, Chris  
; APPLICANT: Wetzel, Jon C.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Vockley, Joseph G.  
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue  
; FILE REFERENCE: 44921-5028  
; CURRENT APPLICATION NUMBER: US/10/240,425  
; CURRENT FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: PCT/US01/09847  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: US 60/193,446  
; PRIOR FILING DATE: 2000-03-31  
; NUMBER OF SEQ ID NOS: 1588  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1585  
; LENGTH: 135259  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. Z83838  
US-10-240-425-1585

Query Match 6.9%; Score 95; DB 6; Length 135259;  
Best Local Similarity 46.4%; Pred. No. 7.6e-15;  
Matches 391; Conservative 0; Mismatches 440; Indels 12; Gaps 2;  
QY 249 CATCTGTCACCCGAGACAGACATGCTTACACTCCATGTCACGATCCCTTCCTCATCTC 308  
DB 69406 CATCATCATCATCTAGTGTGATCATTTACCACCATCACCACCATGATCATCATATATC 69465  
QY 309 CTCTCTGGATCAGACACCAAGTGGTGTCTCAACCTGCGAGTTACACCATCATCTGACTC 368  
DB 69466 CACCGCCACCATCATCTGTGATCAGCACCACCATCACCACCATGATCATC 69525  
QY 369 CGTCAAGACACTCGGTGTGTCATCGGTATCATCTCAACTTCCTGGTTCACCT 428  
DB 69526 CATTTATCACCACCGCCACCATCATCTGTGATCAGCACCACCATCATGATCATCAT 69585  
QY 429 GAGCAGACGAGGAGTCCGAGCAGTCCATCATCTCTCTACGACATCTCCACCAACGC 488  
DB 69586 TATCACCACCATCATCTGTGATCAGCACCACCATCATCTATCACCACCATCATC 69645

```

Qy 489 TCCGTGTTACACAAAGTGGTTCTTCGTACCGGTCCACCAACAAATGANGGTAAATGAA 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69646 CACCACGATCACCAA-----CGTACCACCAACCAACCACTGTGATCAGCACCAT 69696
Qy 549 GATCTACATCAAGCGTAAGTGTATGATGACGACCATCAAGGTCAAGGAGTGAACGGTATCAA 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69697 CACCACCATATCACTACCAACATCAACCAACCAACCAATATCAACCAACCAAC 69756
Qy 609 OTTCTCAGAGACCATCACTTCGAGATCAACAAGATCCAGACACCGGTCGTGATCAGCTC 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69757 CATCACTGTGATCAGCACCATCAACCAACCATATAGCACCATCAACCAACCAACCA 69816
Qy 669 CGATCCGACAAATCAACATGTGGATCCGTGATCTACATCTTCGCAAGAGTGTGA 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69817 CGTCAACCAACCAACCATCACTGATGAGCACCACCATCAACCAACCAACCAATGAT 69876
Qy 729 CGGTAAAGACATCAACATCTGTTCAACTCTTGCAGTACACCAACCGTGTGTCAGAGCTA 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69877 CACCAATATCAACCAACCAACCATCACTGATGAGCACCACCATCAACCAACCAAC 69936
Qy 789 CTGG--GGTAAAGCGTGTATGATGACGAGGAGTACTACATGGTCAACATCGATCTT 845
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69937 CATGATCACCATCTGTCACCAACCAACCATCACTGATGAGCACCACCATCAACCA 69996
Qy 846 GAACAGATATGATGACCAACTCCAGACAGATCGTCTTCAACACAGAGGTAACAA 905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69997 CACCACCAACCATGATCCCATTTATCACCATCACTGTGATCAGCACCATCAACCA 70056
Qy 906 CGACTTCAAGCGGTTACAAAGTATCATCAAGCGTATCAAGGTATCAAGGTAAACCAAG 965
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70057 CACCAATATCAACCAACCAACCATCACTATCAACCAACCAACCAACCAACCAAC 70116
Qy 966 CAGATCAGAGGTTGATCACTCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1025
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70117 CACCATCACCATCATCACCATCATCAACCAACCAACCAACCAACCAACCAAC 70176
Qy 1026 CTTCTTCAAGACAGAGACCATGATGACCGGACACCACTCGACCGGAGGATCAAGC 1085
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70177 CACCACCAACCATCACTGTGATCAGCACCACCATCACTTACCAACCATCAACCA 70236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    1086 CAT 1088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    70237 CAT 70239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
US-09-548-938A-2
; Sequence 2, Application US/09548938A
; GENERAL INFORMATION:
; APPLICANT: EMALFARB, MARK AARON
; APPLICANT: BURLINGAME, RICHARD PAUL
; APPLICANT: OLSON, PHILIP TERRY
; APPLICANT: SINITSYN, ARKADY PANTELEIMONOVICH
; APPLICANT: PARRICHE, MARTINE
; APPLICANT: BOUSSON, JEAN CHRISTOPHE
; APPLICANT: PUNT, PETER JAN
; APPLICANT: VAN-DEIJL, CORNELIA MARIA JOHANNA
; TITLE OF INVENTION: TRANSFORMATION SYSTEM IN THE FIELD OF FILAMENTOUS FUNGI
; FILE REFERENCE: 3123-4001
; CURRENT APPLICATION NUMBER: US/09/548,938A
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3028
; TYPE: DNA
; ORGANISM: Chrysosporium lucknowense
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (46)..(47)
; OTHER INFORMATION: a, t, c, g, other or unknown
; FEATURE:

```

```

; NAME/KEY: modified_base
; LOCATION: (374)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-09-548-938A-2

Query Match      4.2%; Score 58.2; DB 5; Length 3028;
Best Local Similarity 44.3%; Pred. No. 4.4e-06;
Matches 292; Conservative 0; Mismatches 358; Indels 9; Gaps 1;

Qy 364 GACTCGGTCAAGAACAACTCCGGTTGGTCCATCGTATCATCTCCAACTTCTCGTGTTC 423
Db 1729 GCCAAGGGCAAGTCTATCCGCGCCACACCCCTCTCTGSCACTCTCAGTCCGCGAGTGG 1788
Qy 424 ACCTGAGACAGAACAGAGACATCCGAGAGTGCATCAACTTCTCTAGAGATCTCCAAAC 483
Db 1789 GTGAGAAATCAATCAAGACCGCAACACCTTGACCCAGGTCTATCGAGAACCACTCCACC 1848
Qy 484 AAGCTCTCTGTTTACAAAGTGTCTTTCGTACCGTCCACCAACAAATGATGGGTAAAC 543
Db 1849 CTGTCTACTCGTACAAAGGCAAGATCTCTCACTGGGAGCTGTTTAAAGAGATCTTTGCC 1908
Qy 544 ATGAAGATCTACATCAACGGTATGATCGACACCATCAAGGTCAAGGAGTGTGACCGGT 603
Db 1909 GAGGACGGCTCGCTCGGACAGCGCTCTTACGCGCGTCTCGGGGAGGACTTTGTGCGC 1968
Qy 604 ATCAAACTTCTCCAAAGACCATCACTTTCGAGATCAACAAGATCCACAGACACCGGTCTG 663
Db 1969 ATCGCTTCGCGCGCGCGCGCGCGCGCGCGATCCCAAGCGCAAGCTCTACATCAAGACTAC 2028
Qy 664 ACCTCGACATCCGACAAACATCAACATGTGATGATCGGTGATCTTATATCTTTCGCCAAG 723
Db 2029 AACTCGACATTCGCAACTACCGCAAGGTGACCGGGGATGTCGAGAAAGTCAACAAAG 2088
Qy 724 TTGACCGGTAAAGACATCAACATCTCTGTTTCAACTCTCTGACGTACACCAAGTCTGTC 783
Db 2089 TGGATCGCCAGGGGATCCGATCGACGCGATCGGCACCCAGTGCACCTGGCGGGGCC 2148
Qy 784 GACTACTGGGGTA-----ACGACCTGAGATACAAAGAGTACTATGTTCAAC 834
Db 2149 GCGGGGTGGAACACGCGCGCGCGCGCGCGCGCGCTCAAGGCGCTCGCGCGCGCGCAAC 2208
Qy 835 ATCGACTACTTGAACAGATACATGTACGCAACTCCAGACAGATCGTCTTCAACACCAAGA 894
Db 2209 GTCAAGGAGATCGCCATCACCAGCTCGACATCGCGGGCGCTCGCCCAACGACTACCTC 2268
Qy 895 CGTAACAAACAGACTTCAACGAGGTTACAGAGTATCATCAAGCGTATCAGAGGTAAAC 954
Db 2269 ACCGTATGAACGCGCTCGCTCCAGGTCTCCAAAGTGTGCGGATCACCCTCTGGGGCGTC 2328
Qy 955 ACCAAGACACACAGATCAGAGGTGATGATCTCTTACTTCTGACATGACTATCAACAA 1013
Db 2329 TCTGACAAAGCAGCTGGAGGTGAGCAGCAGCAGCGCTCTCTTCTGACAGCACTACCA 2387

RESULT 3
US-10-092-411A-1515/c
; Sequence 1515, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 1515
; LENGTH: 2835

```

TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-1515

Query Match 3.7%; Score 50.6; DB 6; Length 2835;  
Best Local Similarity 41.8%; Pred. No. 0.00041;  
Matches 378; Conservative 0; Mismatches 524; Indels 3; Gaps 1;

Qy	286	ATGATACAGTCCCTTCTCCATCTCTCTGATCAGATCAAAAGGCGGTCTCCAACTTG	345
Db	2654	ATCAGCGAGTACGTCGACATCTGATAGTCATCAGAGTCAAGAAAGAGTGAATCAAC	2595
Qy	346	CCAGGTATACATCATCGACTCCGTCAGAAACAATCCCGTGGTTCATCGGTATCATC	405
Db	2594	GAGTAAAGTACGTCATTAAGTGAATCGAAGCAAGAGCTTTCAGATTCAGCAAGCGT	2535
Qy	406	TCCAACTTCCTGGTCTTCCACCTCGAGCAAGAGGAGTCCGAGCAGTCCATCAACTTC	465
Db	2534	GTCAACGTCAAGAGTGGTTCACATCAACAGTGTGAGTGGCTCAACAAGTACAGCAT	2475
Qy	466	TCTTACGACATCTCCAAACAGCTCTCGGTTCACAAAGTGTCTTCTCCACCGTCCAC	525
Db	2474	TTCAGACTCAACGAGCATCGAGCTGTGATAGCGCATCAATAAAGCGAGGGAATCAGC	2415
Qy	526	AACACATGATGGTACATGAGATCTACATCAACGTTAAGTGTATCCGACCATCAG	585
Db	2414	AAGCACAAGTAACTATTAAGTAACTGATGAGTGTGATAGCGCATCAATAAAGCGAGG	2355
Qy	586	GTCAAGAGTGTACCGGTATCAACTTCTCCAGACCATCATCCTTCGAGATCAACAAGATC	645
Db	2354	GTCAAGAGGTAACTGATCTCCATCGCAAGTACGTCTATTAAGTAAATCGCAAGCAC	2295
Qy	646	CCAGACCGGTCTGATCACTCCGACTCCGACCAACATCAACATGTGGATCCGTGACTTC	705
Db	2294	AAGCGTTTCAGATTCAACAAGTACATCGACATCCGACAGTGGCTCCACATCAACGAGTGA	2235
Qy	706	TACATCTTCGCAAGAGTGTGGAGGTAAAGACATCAACATCTGTTCAACTCTTTCGAG	765
Db	2234	GAGTGAGTCAGACAGTGCAGTACGTATTAAGTGAATCGACAAGTACAAGGTTTCAGA	2175
Qy	766	TACACCAAGTCTGATGAGTCTCTGGGTAAAGCAAGCTTGATGATCAACAAGAGTACTAC	825
Db	2174	TTCACGAGTACGTGACATCCGACATCGATGATCAATGTCAAGCGAGGAGTGAATCAAA	2115
Qy	826	ATGGTCAACATCGACTACTTGAACAGATACATGATAGCCCAACTCCAGACAGATCGCTTC	885
Db	2114	CAGTAAAGTACGTCATTAAGTGAATCGACAAGTACGAGTCTTTCAGGCTCAACGAGTGC	2055
Qy	886	AACACCAAGTAAACAACAGCTTCAACGAGGTTTCAAGATCATCATCAAGCGTATC	945
Db	2054	GTGACATCCGACAGTGCATCAAGCTCAACGAGTGAAGTGAAGTACATCGTTGAGTGA	1995
Qy	946	AGAGTAAACACCAACGACACACAGATCAGAGTGGTGACATCTCTGATTCGACATGACT	1005
Db	1994	GTCAATTAAGTAAATCGACAGTACGAGTCTTTCAGGCTCAACGAGTGGTGCATCCGA	1935
Qy	1006	ATCAACACAGGCTTCAACCTCTTCAAGAAAGAGACCATGATACGCGGACACACAC	1065
Db	1934	CAGTGCATCAACGTCACAAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA	1875
Qy	1066	TCCACCGAGGACATCTCGCCATCGGCTCGGTGACGACAGCAAGGACATCAACGACAC	1125
Db	1874	ATCATTAAGGACA---AGCGTTTCAGATTCACAAAGCGCGTCAACGTCAGAAAGTGCATC	1818
Qy	1126	ATCATCTTCCAGATCCAGCAATGAACAACATTAATTAATTAATTAATTAATTAATTAAT	1185
Db	1817	CACATCAACGAGTCAAGATGAGTCAACACAGTCCGACGACATCGTTAAGTGGATCATTAAG	1758
Qy	1186	TCCAA 1190	
Db	1757	TACAA 1753	

RESULT 4  
US-10-092-411A-1685/c  
Sequence 1685; Application US/10092411A  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-101  
CURRENT APPLICATION NUMBER: US/10/092,411A  
CURRENT FILING DATE: 2002-03-07  
PRIOR APPLICATION NUMBER: US 09/134,001  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5676  
SEQ ID NO 1685  
LENGTH: 3561  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-1685

Query Match 3.7%; Score 50.6; DB 6; Length 3561;  
Best Local Similarity 41.8%; Pred. No. 0.00046;  
Matches 378; Conservative 0; Mismatches 524; Indels 3; Gaps 1;

Qy	286	ATGATACAGTCCCTTCTCCATCTCTCTGATCAGATCAAAAGTGGTCTCCAACTTG	345
Db	3450	ATCAGCGAGTACGTCGACATCTGATAGTGCATCAGATCAGCAAGAAAGAGTGAATCAAC	3391
Qy	346	CCAGGTATACATCATCGACTCCGTCAGAAACAATCCCGTGGTTCCTCGGTATCATC	405
Db	3390	GAGTAAAGTACGTCATTAAGTGAATCGACCAAGGTTTCAGATTCAGCAAGCGT	3331
Qy	406	TCCAACTTCCTGGTCTTCCACCTCGAGCAAGAGGAGTCCGAGCAGTCCATCAACTTC	465
Db	3330	GTCAACGTCAAGAAAGTCCGTCACATCAACAGTGTGAGTGGCTCAACAAGTACAAGCAT	3271
Qy	466	TCTTACGACATCTCCAAACAGCTCTCGGTTCACAAAGTGTCTTCTCGTACCGTCCAC	525
Db	3270	TTCAGATCAACGAGCACATCGAGCTGTGATAGCGCATCAATAAAGCGAGGAAATCAGC	3211
Qy	526	AACACATGATGGTAAACATGAAGATCTACATCAACGAGTGAAGTGAAGTGAAGTGAAGT	585
Db	3210	AAGCACAAGTAACTATTAAGTGAATCGACAAGTACGTCGAGTGTGATAGCGCATCAAC	3151
Qy	586	GTCAAGAGTGTGCGGTATCAACTTCTCCAGACCATCACTTCGAGATCAACAAGATC	645
Db	3150	GTCAACGAGCGTAACTGACTCCCAATAGCGCAAGTACGTCATTAAGTAAATCCAGAGCAC	3091
Qy	646	CCAGACCGGTCTGATCACTCCGACTCCGACCAACATCAACATGTGGATCGGTGACTTC	705
Db	3090	AAGCGTTTCAGATTCAACAAGTACATCGACATCCGACAGTGGCTCCACATCAACGAGTGA	3031
Qy	706	TACATCTTCGCAAGAGTGTGAGGATGAGGATCAACATCTCTGATTCCTTCACTTTCAG	765
Db	3030	GAGTGAAGTCAAGAGTCCCAAGTACGTCATTAAGTGAATCGACAAGTGAAGGTTTCAGA	2971
Qy	766	TACACCAAGTCCGTCAGAGTACTTCTGGGTGAAGGAGTGAAGTGAAGTGAAGTGAAGT	825
Db	2970	TTCACAGAGTACGTCGACATCCGACAGTGCATCAATGTCAAGGAGGAGTGAATCAAA	2911
Qy	826	ATGTCAACATCGACTACTTGAACAGATACATGTAGCCCACTCCAGACAGATCGCTTTC	885
Db	2910	CAGTAAAGTACGTCATTAAGTGAATCGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA	2851
Qy	886	AACACCAAGGATCAACAACAGCTTCAACGAGGTTTACAAGATCATCATCAAGGATTC	945
Db	2850	GTCCACATCCGACAGTGCATCAACGTCACAGGTGAGTGAAGTGAAGTGAAGTGAAGTGA	2791
Qy	946	AGAGTAAACACCAACGACACACAGTCAAGGTTGATCACTGTGACTTCGACATGACT	1005

Db 2790 GTCATTAAAGTGAATCGACAAAGTACGAGTCTTTCAGGCTCAAGAGTGCCTCGACATCCGA 2731  
Qy 1006 ATCAACAAAGGCTTCAACAGTCTTTCATGAAGAGGAGGACCATGTACCGCCGACAAACAC 1065  
Db 2730 CAGTGCATCAAGCTCAACAGAGGAGTACTCAACAAGTGAAGTACATCGTTGATGA 2671  
Qy 1066 TCCACGAGAGACATCTACGCGCATCGGTCTCGGTGAGACAGACAGACATCAACAGCAAC 1125  
Db 2670 ATCATTAAAGCACA---AGCGTTTTCAGATTCAACAAGCGGTCAACGTCAGAAAGTGCATC 2614  
Qy 1126 ATCATCTCCAGATCCAGCCCAATGAACAACACTTACTAGCGTTCCCGAGATCTTCAAG 1185  
Db 2613 CACATCAACGAGTCAAGTCAACAGTGCACACATCGTTAAAGTGGATTAAG 2554  
Qy 1186 TCCAA 1190  
2553 TACAA 2549  
RESULT 5  
US-10-092-411A-1626  
Sequence 1626, Application US/10092411A  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-101  
CURRENT APPLICATION NUMBER: US/10/092,411A  
CURRENT FILING DATE: 2002-03-07  
PRIOR APPLICATION NUMBER: US 09/134,001  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5676  
SEQ ID NO 1626  
LENGTH: 6414  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-1626  
Query Match 3.7%; Score 50.6; DB 6; Length 6414;  
Best Local Similarity 41.8%; Pred. No. 0.00062; Mismatches 524; Indels 3; Gaps 1;  
Matches 378; Conservative 0;  
286 ATGTACGAGTCCCTTCTCCATCTCTCTGATCAGATCAACAAGTGGTCTCCCACTTG 345  
2595 ATCAGCGAGTACGTCGACATCTGATAGTGCATCAGATCAGCAAGAAAGTGAAGTCAAC 2654  
346 CCAGGTACACCATCATCGACTCGGTCAAGAACTCCGGTTGGTCCATCGGTATCATC 405  
2655 GAGTAAAGTACGTCATTAAGTGAATCGAAGCAAGCGTTTCAGATTCAGACAGCGT 2714  
406 TCCAACTCTCTGGTCTTCAACCTGAAGCAGAGGAGTCCGAGCAGTCCCATCACTTC 465  
2715 GTCAAGTCAAGAAAGTGGTCAACATCAACAGTGTGAGTGGCTCAACAAGTACAGCAT 2774  
466 TCCTAGACATCTCCACAGCGTCTCTGGTTTACAACAAGTGGTCTTGTACCGGTCAAC 525  
2775 TCCAGATCAACAGGACATCGAGCTGTGATCGCATCAATAAAGCGGAGGAATCAGC 2834  
526 AACAACTATGATGGTGAATCAAGATCTACATCAAGGTAAGCTGATCGACACCATCAAG 585  
2835 AAGCAAGTAAACTATATCAGNATCAGTGAAGTACGTCGACGCTGTAGCGCATCAAC 2894  
586 GTCAGAGGTGACCGGTATCACTTCTCCAGACATCACTTCGAGATCAACAGATC 645  
2895 GTCAGAGGTGACCGGTATCACTTCTCCAGACATCACTTCGAGATCAACAGATC 645  
546 CCAGACCGGTCTGATCACTCGAGTCCGACAAACATCAACATCACTGATGGATCGGTGACTTC 705  
2955 AAGCGTTTCAAGTCAACAGTATCACTCGACATCGGACATCGGTCGATCGATCAACAGTGA 3014

Qy 706 TACATCTTCCCAAGAGGTGGACGGTAAAGGACATCAACATCTCTTCAACTCCCTTCGAG 765  
Db 3015 GAGTGAAGTCAAGAGTCAAGTACGTCATTAAGTGAATCGAAGTCAAGCGTTTCAAG 3074  
Qy 766 TACACCAAGCTCTCTCAAGGACTACTGGGTAAAGGCTTGAAGATCAACAGAGGATCTAC 825  
Db 3075 TTAACAGGATCTCTCAAGGACTACTGGGTAAAGGCTTGAAGATCAACAGAGGATCTAC 825  
Qy 826 ATGTCACATCTCAAGGACTACTGGGTAAAGGCTTGAAGATCAACAGAGGATCTAC 885  
Db 3135 CAGTAAAGTACGTCATTAAGTGAATCGAAGTCAAGGCTTTCAGGCTCAACGAGTGC 3194  
Qy 886 AACACAGAGCTTCAACACAGGACTTCAACAGGAGTTCACAGATCATCATCAAGGATTC 945  
Db 3195 GTCAGATCTCAAGGACTTCAACAGGAGTTCACAGATCATCATCAAGGATTC 945  
Qy 946 AGAGTAAACACCAAGGACTTCAACAGGAGTTCACAGATCATCATCAAGGATTC 1005  
Db 3255 GTCATTAAGTGAATCGACAGTACGAGTCTTTCAGGCTCAACGAGTTCGACATCCGA 3314  
Qy 1006 ATCAACAAAGGCTTCAACAGGACTTCAACAGGAGTTCACAGATCATCATCAAGGATTC 1065  
Db 3315 CAGTGAAGTCAAGGACTTCAACAGGAGTTCACAGATCATCATCAAGGATTC 1065  
Qy 1066 TCCACCAAGGACTTCAACAGGACTTCAACAGGAGTTCACAGATCATCATCAAGGATTC 1125  
Db 3375 ATCATTAAGTGAATCGACAGTACGAGTCTTTCAGGCTCAACGAGTTCGACATCCGA 3431  
Qy 1126 ATCATCTTCCAGATCCAGGACTTCAACAGGAGTTCACAGATCATCATCAAGGATTC 1185  
Db 3432 CACATCAACGAGTTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3491  
Qy 1186 TCCAA 1190  
Db 3492 TACAA 3496

## RESULT 6

US-10-085-198-111  
Sequence 111, Application US/10085198  
GENERAL INFORMATION:  
APPLICANT: Alsobrook et al  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-279  
CURRENT APPLICATION NUMBER: US/10/085,198  
CURRENT FILING DATE: 2002-02-25  
PRIOR APPLICATION NUMBER: 60/271,646  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 60/276,401  
PRIOR FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/311,981  
PRIOR FILING DATE: 2001-08-13  
PRIOR APPLICATION NUMBER: 60/312,858  
PRIOR FILING DATE: 2001-08-16  
PRIOR APPLICATION NUMBER: 60/271,840  
PRIOR FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: 60/277,324  
PRIOR FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: 60/286,096  
PRIOR FILING DATE: 2001-04-21  
PRIOR APPLICATION NUMBER: 60/299,695  
PRIOR FILING DATE: 2001-06-20  
PRIOR APPLICATION NUMBER: 60/315,614  
PRIOR FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/272,405  
PRIOR FILING DATE: 2001-02-28  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 653  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 111  
LENGTH: 15645  
TYPE: DNA

ORGANISM: Homo sapiens  
US-10-085-198-111

Query Match 3.7%; Score 50.4; DB 6; Length 15645;  
Best Local Similarity 47.0%; Pred. No. 0.0011;  
Matches 190; Conservative 0; Mismatches 211; Indels 3; Gaps 1;

QY 579 CATCAAGGTCAGAGGTTGACCGGTATCAACTTCCTCAAGACCATCACTTCGAGATCAA 638  
DB 9784 CAGCAGGGTCAGGGCCTTATGGACACAGGCTTGGTCACAGCCCGCAGCAGCAGCAA 9843  
QY 639 CAAGATCCAGACACAC---GGTGTATCACTCCGACTCCGACACATCAATGTGGAT 695  
DB 9844 CAACAGCACCAAGCAGAGGGTCCATGGCAGGGCTGTCCATCTTCAGCAAGTCTGATG 9903  
QY 696 CCGTGACTTTCATCTTCGCCAAGGAGTTGGACGGTAAAGGATCAACATCCTCTTCAA 755  
DB 9904 TCACAGAGTGGGCGAGCCCAACTGAGCGCTCAGCCCATGGGCTCTTACAGCAGGTTGAG 9963  
QY 756 CTCCTTGCAGTACACCAAGTCGTCAGGAGCTACTGCGGTACGACCTGAGATACACAA 815  
DB 9964 CAGCAGCAGCGTGCACAGCAAGCAAGCAACTTCAGCAGCAGCAGCAGCAGCTACAA 10023  
QY 816 GGAGTACTATGTTCAACATCGACTTACTTGAACAGATACATGTACGCCAACTCCAGACA 875  
DB 10024 CAGCAACAGCAACTTCAGCAGCAACAGCTTTCACAGCAGCAACAGCAGCAGCTTCAA 10083  
QY 876 GATCGCTTTCACACACAGCAGTTAAACCAACGACTTCAACGAGGTTTACAGATCATCAT 935  
DB 10084 CAACAGCAGCAGCAACAGCTTCAACAGCAGCAACAGCAGCTTCAACAGCAGCAACAA 10143  
QY 936 CAGCGGTATCAGAGGTAAACCAACAGCAACAGCAGCAGTACAGAGTG 979  
DB 10144 CACAGCAGCAGTTTCAACAGCAGCAGCAGCAGCAGCAGCAGTGGG 10187

## RESULT 7

US-10-270-223-1  
Sequence 1, Application US/10270223  
GENERAL INFORMATION:  
APPLICANT: BioImage A/S  
TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS  
TITLE OF INVENTION: IMPACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION FROM  
TITLE OF INVENTION: INTERACTIONS BY FLUORESCENCE REDISTRIBUTION.  
FILE REFERENCE: 3759-0126P  
CURRENT APPLICATION NUMBER: US/10/270,223  
CURRENT FILING DATE: 2002-10-11  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1

LENGTH: 1452  
TYPE: DNA  
ORGANISM: Aequoria Victoria and Human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1452)  
OTHER INFORMATION:  
US-10-270-223-1

Query Match 3.5%; Score 48.4; DB 6; Length 1452;  
Best Local Similarity 45.2%; Pred. No. 0.0011;  
Matches 218; Conservative 0; Mismatches 261; Indels 3; Gaps 1;

QY 624 CACCTTCGAGATCAACAGATCCAGACACCGTGTGATCACTCCGACTCCGACAT 683  
DB 177 CACCCCTGTGACACCCCTGACCTGCGGTGCGTGTCTTACGCGCTTACCCGACCAT 236  
QY 684 CAACATGTGGATCCGTGACTTCTACATCTTCGCCAAGGAGTTGGACGTTAAGGACATCAA 743  
DB 237 GAGCAGCAGCAGCTTCTTCAAGTCCGCCATGCCGAGGCTACGTTCCAGGAGCGCACCAT 296  
QY 744 CATCTGTTCACCTCTTCAGTACACCAACGTCGTCAAGGAGTACTTGGGTGAGGAGCT 803

DB 297 CTTCTTCAAGGACGAGCGCAACTACAAGACCGCGCGCGAGGTGAAGTTTCGAGGGCGACAC 356  
QY 804 GAGATCAACAAGGAGTACTACATGTTCAACATCGACTACTTGAACAGATACATGTACGC 863  
DB 357 CTGTGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGCGCAACATCCITGGG 416  
QY 864 CAACCTCCAGCAGATCGTCTTCAACACACAGAGGTAAACCAACAGCAGCTTCAACAGGAGTTA 923  
DB 417 GCA---CAAGCTGGAGTACAACATACAACAGCCACACGCTCTATATCATGCGCGACAGCA 473  
QY 924 CAAGATCATCATCAAGCGTATCAGAGGTAAACCAACAGCAGCAGAGTCAGAGTGTGTA 983  
DB 474 GAGAGAGGCTATCAGGTGAATTCAGATCCGCCACACATCGAGGACGCGAGGTGCA 533  
QY 984 CATCTGTACTTTCGACATGACTATCAACAAAGCCCTACAACTGTTTCATGAAGAACA 1043  
DB 534 GCTCGCGGACCACTACACAGCAGAACACCCCATCGCGACGCGCCGCTGCTGCTGCCGA 593  
QY 1044 GACATGTAGCGGACAAACCACTCCACCGAGGACATCTACGCCATCGTCTGCGTGAGCA 1103  
DB 594 CAACCTACTCTGAGCAGCCAGTCCGCTGAGCAAGAGACCCCAACAGAGACGCGGTCA 653  
QY 1104 GA 1105  
DB 654 CA 655

## RESULT 8

US-10-092-411A-1670/c  
Sequence 1670, Application US/10092411A  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-101  
CURRENT APPLICATION NUMBER: US/10/092,411A  
CURRENT FILING DATE: 2002-03-07  
PRIOR APPLICATION NUMBER: US 09/134,001  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5676  
SEQ ID NO 1670  
LENGTH: 783  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-1670

Query Match 3.5%; Score 47.6; DB 6; Length 783;  
Best Local Similarity 42.8%; Pred. No. 0.0013;  
Matches 239; Conservative 0; Mismatches 319; Indels 0; Gaps 0;

QY 521 TCACCAACACATGATGGTAAACATCACTACATCAACGGTAAAGTGTATCGACACCA 580  
DB 610 TCAGATTCAACAGTACATCGACATCCGACAGTGGTCAACGTCACCAAGTGAGAGTGAC 551  
QY 581 TCAAGTCAAGGAGTTGACCGGTATCAACTTCTCAAGACCATCACTTCGAGATCAACA 640  
DB 550 TCAACAGTGAAGTACATCGTTAAGTGTAGTCGACAGCAAGCGTTTCAGATTCAACA 491  
QY 641 AGATCCACACACCGGTGTGATCACTCCGACTCCGACCAACATCAACATGTGGATCGTG 700  
DB 490 AGGCGTCAACGTTCAGAAAGTGCATCAACATCAACGAGCGTGAGTGACTTCCAAATAGCGCA 431  
QY 701 ACTTCTACATCTTCGCCAAGGAGTTGGACGGTAAAGGATCAACATCTCTTCAACTCTT 760  
DB 430 AGTACCTCATTAAGTGGCTCAACAGTCAACAGCGTTTCAGAGTCAACGAGTACGTCGACA 371  
QY 761 TGCAGTACACCAAGTGTCAAGGACTACTGGGTAAAGACCTGAGATACACAGGAGT 820  
DB 370 TCAGCAAGTGCATCAACATCAACGAGCGGAGCGGACTCCGATAGCGCAAGTACGCTGTTA 311



QY 821 ACTATGGTCAACATCGACTACTTGAACAGATACATGTACGCCACTCCAGACAGATCG 880  
 DB 310 AGTGGCTCAACAAGTCAAGCATTTTCAGACTCAACAGTACGTGCGACATCAGACAGTGGG 251  
 QY 881 TCTTCACACAGAGGTAAACAACAGACTTCAACGAGGTTTACAGATCATTCAACG 940  
 DB 250 TCCATATCCATCAGAAAGTGCATCCATCAACGAGTGTGAGTGAAGTCAACAGTGAA 191  
 QY 941 GTATCAGAGTAAACCAACGACACACAGAGTCAAGAGTGGTGATCTTCTACTTTCGACA 1000  
 DB 190 AGTATATCGGTAAAGTCAATCAAGTACAAAGGTTTCAGATTCACAAAGTACATCGACA 131  
 QY 1001 TGATATCAACAACAGGCTTACAACTGTTTCATGAGACGAGACCATGTACGCGGACA 1060  
 DB 130 TCAGAAAGTGTCAAGCTCAACAGCGAGAGTGAATCAACAAGTGAAGTACATCGGTA 71  
 1061 ACCACTCCACCGAGGACA 1078  
 70 AGTGAATCATCAAGTACA 53

RESULT 9  
 Sequence 138, Application US/10240965  
 GENERAL INFORMATION:  
 APPLICANT: INCYTE GENOMICS, INC.  
 APPLICANT: SHIFFMAN, Dov  
 APPLICANT: SOMOGYI, Roland  
 APPLICANT: LAWN, Richard M.  
 APPLICANT: SEILHAWER, Jeffrey J.  
 APPLICANT: PORTER, Gordon J.  
 APPLICANT: MIKITA, Thomas  
 APPLICANT: TAL, Julie  
 TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION  
 FILE REFERENCE: PA-0025 PCT  
 CURRENT APPLICATION NUMBER: US/10240,965  
 CURRENT FILING DATE: 2002-10-04  
 PRIOR APPLICATION NUMBER: 60/195,106  
 PRIOR FILING DATE: 2000-04-05  
 NUMBER OF SEQ ID NOS: 276  
 SOFTWARE: PERL Program  
 SEQ ID NO 138  
 LENGTH: 8249  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 NAME/KEY: misc\_feature  
 FEATURE:  
 OTHER INFORMATION: Incyte ID No: 445012.6  
 NAME/KEY: unsure  
 LOCATION: 4150-4244, 8100-8103, 8105, 8107, 8109-8110, 8115, 8120, 8127, 8131,  
 LOCATION: 8137, 8150, 8155, 8157, 8180  
 OTHER INFORMATION: a, t, c, g, or other  
 US-10-240-965-138

Query Match 3.4%; Score 46.8; DB 6; Length 8249;  
 Best Local Similarity 41.9%; Pred. No. 0.0069;  
 Matches 343; Conservative 0; Mismatches 472; Indels 3; Gaps 1;

QY 251 TCGTACCCAGAGAGACATCTCTACACTCCATGTACGAGTCTCTTCATCTCT 310  
 DB 1762 TCCACAAACAGTCCGTTGTGCGGAGAGATGGCTTTGAACAAGAAGTTGCGTGTGAGC 1821  
 QY 311 TGTGATCAGATCAACAAGTGGGTCTCCAACTTGGCGGCTTACACCATCATCGACTCGG 370  
 DB 1822 ATGGCATTTCCACAGACATGGGTATGAGTGGCGGCCCGCCACTCGGGGTGTACCCCG 1881  
 QY 371 TCAGAGAAATCCCGTTGGTTCATCGGTATCTCCAACTTCTCTGGTCTTCCACCTGA 430  
 DB 1882 TGCAGTCACTGTACGAGCTTGAAGCAGGTGTGGAGCATCCGCGTGACCAAGACGG 1941  
 QY 431 AGCAGACGAGGACTCCGAGGAGTCCATCACTTCTCTACGACATCTCCAAACAGCTC 490

DB 1942 AGGAGTACCCCGCCAGCCGCTACCGCGTTCATCCACAATGCGATCA 2001  
 QY 491 CTGTTTACAACAAGTGGTTCTTCGTCAACGCTCACCACAACATGATGGTAAACATGAAGA 550  
 DB 2002 TGGTTCTCCACGCGAGACGCTGCGGCTCTTCCACACACACCATCTTCTACACAGGTACC 2061  
 QY 551 TCTACATCAAGGTAAAGTGTATCGACACCATCAAGTCAAGGTGACCGGTATCACT 610  
 DB 2062 CTGCGGCTCCAGTGAAGTCAAGATCATCAACGGGGCGAGGTCTTCTCCACCGTGC 2121  
 QY 611 TCTCAAGACATCACTTCCAGATCAACAAGATCCAGACACCGCTGTGATCACTCGG 670  
 DB 2122 TCTCATCTCTATCAGATCTTCTATGCGACCTGTCCAACTATGGAATGACCGCTGG 2181  
 QY 671 ACTCCGACATCAACATCTGGATCGGTGACTTCTACATCTTCCCAAGGAGTGGAGG 730  
 DB 2182 GCTGTATACCTTCAAGCACCTGTGGCTTCTGACTCTCTGGACCAACCTCCGCGTGC 2241  
 QY 731 GTAAGGACATCAACATCTCTTCAACTCTTTCAGTACACCAACGCTGTCAGAGGACTACT 790  
 DB 2242 AGACATGCCCCCTGTGACTTG---GCGCAGAAGTACTTCCAGATCTTCTCCGAGGAGA 2298  
 QY 791 GGGTAACGACCTGAGATACACAAGAGTACTACATGTGTCACATCTGACTACTTGACACA 850  
 DB 2299 AGGACCGGCTCTGCGAGGACCCCTGCGAGGACAAACGCTCACAAGACATCTGGTCCAAAG 2358  
 QY 851 GATACATGTAGCCCACTCCAGACAGATCGTCTTCAACACACAGACGTAAACAACAACGACT 910  
 DB 2359 AGAAGAGTGTGACCGCTTCCCAAGCTCTCATCATCGGCCCCAGAAACAGGACCA 2418  
 QY 911 TCAACGAGGTTTAAAGATCATCATCAAGCGTATCAGAGTATACACCAACGACACAGAG 970  
 DB 2419 CTGCGCTCTACTCTTCTGCGGATGACCTGACCTTAAGCAGCAACTACCCGAGTCTG 2478  
 QY 971 TCAGAGTGTGACATCTCTGACTTCGACATGACTATCAACACAAGGCTACACCTGT 1030  
 DB 2479 AGACTTTGAGAGATCCAGTTTATTTATGGCCACAACTATCAAAAGGCATCGACTGT 2538  
 QY 1031 TCATGAAGACGAGACCATGTACGCCGACAAACCACTCC 1068  
 DB 2539 ACATGGAGTTCTTCCCATCCCTTCCCAACACCACTCC 2576

RESULT 10  
 US-10-267-255-83  
 ; Sequence 83, Application US/10267255  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D  
 ; APPLICANT: Mao, Y  
 ; APPLICANT: Varoglu, M  
 ; APPLICANT: He, M  
 ; APPLICANT: Sheldon, P  
 ; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
 ; FILE REFERENCE: 600.456US1  
 ; CURRENT APPLICATION NUMBER: US/10/267,255  
 ; CURRENT FILING DATE: 2002-10-09  
 ; PRIOR APPLICATION NUMBER: US 09/266,965  
 ; PRIOR FILING DATE: 1999-03-12  
 ; PRIOR APPLICATION NUMBER: US 08/624,447  
 ; PRIOR FILING DATE: 1996-08-19  
 ; PRIOR APPLICATION NUMBER: PCT/US94/11279  
 ; PRIOR FILING DATE: 1994-10-06  
 ; PRIOR APPLICATION NUMBER: US 08/133,963  
 ; PRIOR FILING DATE: 1993-10-07  
 ; NUMBER OF SEQ ID NOS: 145  
 ; SOFTWARE: PastSeq for Windows Version 3.0  
 ; SEQ ID NO 83  
 ; LENGTH: 2466  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces lavendulae  
 US-10-267-255-83

```

Query Match      3.4%; Score 46; DB 6; Length 2466;
Best Local Similarity 43.2%; Pred. No. 0.0061;
Matches 269; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

QY 470 ACACATCTCCACACAGCTCCCTGGTTAGAACAAAGTGTCTTCGTACCCGTCACCAACA 529
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1517 ACAGCAAGCGCGACATCGCGGTCTGTACGACATGGGTGCGACTGAGGACGCGCGCAACC 1576
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 530 ACATGATGGTAAATGAAGATACATCAAGGGTAAGCTGATCGACACCATCAAGGTCA 589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1577 GCACCAAGTTGTACGAGTTCCACGACCGCGAGCGGATTAACAGCGCGGTCAAGGTCT 1636
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 590 AGAGTTGACCGGTATCACTTCTCCACAGCATCACTTCAGATCAACAGATCCAG 649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1637 GGGACAGACAGCAGCGCGGTCAAGAGCTGGAAGTGGGCTCCAGCAAGTGAACGTCG 1696
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 650 ACACCGGTCTGATCACTCGACTCCGACCAACATCAACATGATGGTACTTCTACA 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1697 GCGACTTCGACGCGGAGCGCAAGCGCGACATCGCGTCTGTACGACTACGCGAGGAG 1756
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 710 TCTTCGCCAAGGAGTTGGAGGTAAGGATCAATCACTCTGTTCACCTCTTGCAGTACA 769
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1757 GCGACCGCAGCGGTACGGGCTGTGGACCTTCACGACGCGGTGCGGCTTCAACCGGCC 1816
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 770 CCAAGCTCGTCAAGGACTACTGGGGTAAGACCTGAGATACAAAGGAGTACTACATGG 829
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1817 CCAAGCTGTGTGGACAGCAAC--AACGACCGGTCAAGAGCTGGAAGTGAAGCGCA 1873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 830 TCAATCTGACTACTTGAACAGATACATGTACGCCCACTCCAGACAGATGCTCTTCAACA 889
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1874 GCAAGCCACCGTCGGGACTTCAAGCGGAGCGCAAGCGGACATCGGCTCTCTACG 1933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 890 CCAGACTTAAACAAGGACTTCAAGAGGGTTACAGAGATCATCAAGCGTATCAGAG 949
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1934 ACATGSGTCGACGAGGAGCGCGCAACCGCACCAAGCTGTTCACCTTACCGCGCAGG 1993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 950 GTAACCAACAGCACACAGATCAGAGGTGGTACATCTCTGTTCGACATGACTATCA 1009
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1994 CGACCGGTTTCAACAGCGCGGTCAAGGTGTGGACAGCAACGACGACCGGTGAAGAGCT 2053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1010 ACAACAGGCTTCAACCTCTTCAAGAGAGGATGACAGACGACCATGACGCGCAACCACTCCA 1069
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2054 GGAAGTGGAGCGCGTCAAGGTAGTGGAGGCGACTTCAACGCGGACGCAAGCGGACA 2113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1070 CCGAGGACATCTAGCCATCGG 1091
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2114 TCGGGGTGTGTACGACTACGG 2135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-10-267-255-75
Sequence 75, Application US/10267255
GENERAL INFORMATION:
APPLICANT: Sherman, D
APPLICANT: Mao, Y
APPLICANT: Varoglu, M
APPLICANT: He, M
APPLICANT: Sheldon, P
TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
FILE REFERENCE: 600.456US1
CURRENT APPLICATION NUMBER: US/10/267,255
CURRENT FILING DATE: 2002-10-09
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 09/266,965
PRIOR FILING DATE: 1996-08-19
PRIOR APPLICATION NUMBER: PCT/US94/11279
PRIOR FILING DATE: 1994-10-06
PRIOR APPLICATION NUMBER: US 08/133,963
PRIOR FILING DATE: 1993-10-07
NUMBER OF SEQ ID NOS: 145
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 75

```

```

; LENGTH: 18034
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
; NAME/KEY: unsure
; LOCATION: (302)...(302)
; OTHER INFORMATION: n is a or t or g or c
US-10-267-255-75

Query Match      3.4%; Score 46; DB 6; Length 18034;
Best Local Similarity 43.2%; Pred. No. 0.017;
Matches 269; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

QY 470 ACACATCTCCACACAGCTCCCTGGTTAGAACAAAGTGTCTTCGTACCCGTCACCAACA 529
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13031 ACAGCAAGCGCGACATCGCGGTCTGTACGACATGGGTGCGACTGAGGACGCGCGCAACC 13090
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 530 ACATGATGGTAAATGAAGATACATCAAGGGTAAGCTGATCGACACCATCAAGGTCA 589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13091 GCACCAAGTTGTACGAGTTCCACGACCGCGAGCGGATTAACAGCGCGGTCAAGGTCT 13150
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 590 AGAGTTGACCGGTATCACTTCTCCAGACCATCACTTCGAGATCAACAGATCCAG 649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13151 GGGACAGACAGCAGCGCGGTCAAGAGCTGGAAGTGGGCTTCCAGCAAGTGAACGTCG 13210
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 650 ACACCGGTCTGATCACTCGACTCCGACCAACATCAACATGTGGATCCGTACTTCTACA 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13211 GCGACTTCGACGCGGAGCGCAAGCGCGACATCGCGTCTGTACGACTACGCGAGGAGG 13270
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 710 TCTTCGCCAAGGAGTTGGAGGTAAGGATCAACATCTCTGTTCACCTCTTGCAGTACA 769
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13271 GCGACCGCAGCGGTACGGGCTGTGGACCTTCCACGACGACCGGTGCGGCTTCAACCGGCC 13330
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 770 CCAAGCTCGTCAAGGACTACTGGGGTAAGACCTGAGATACAAAGGAGTACTACATGG 829
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13331 CCAAGCTGTGTGGACAGCAAC--AACGACCGGTCAAGAGCTGGAAGTGAAGCGCA 13387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 830 TCAACATCTGACTACTTGAACAGATACATGTACGCCCACTCCAGACAGATGCTCTTCAACA 889
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13388 GCAAGCCACCGTCGGGACTTCAAGCGGAGCGCAAGCGGACATCGGCTCTCTACG 13447
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 890 CCAGAGTAAACAAGGACTTCAACGAGGGTTACAAGATCATCAAGCGTATCAGAG 949
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13448 ACATGGTCCACCGAGGAGCGCGCAACCGCAACAGCTGTTCACCTTACCGCGCAGG 13507
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 950 GTAACCAACAGCACACAGATCAGAGGTGGTACATCTCTGTTCGACATGACTATCA 1009
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13508 CGACCGGTTTCAACAGCGCGGTCAAGGTGTGGACAGCAACGACGACCGGTGAAGAGCT 13567
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1010 ACAACAGGCTTCAACCTCTTCAAGAGAGGATGACAGACGACCATGACGCGCAACCACTCCA 1069
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13568 GGAAGTGGAGCGCGGTCAAGGTAGTGGGAGGCGACTTCAACGCGGACGCAAGCGGACA 13627
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1070 CCGAGGACATCTAGCCATCGG 1091
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 13628 TCGGGGTGTGTACGACTACGG 13649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-10-270-223-5
Sequence 5, Application US/10270223
GENERAL INFORMATION:
APPLICANT: Bioimage A/S
TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR CO
TITLE OF INVENTION: INTACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATIO
FILE REFERENCE: 3759-0126P
CURRENT APPLICATION NUMBER: US/10/270,223
CURRENT FILING DATE: 2002-10-11
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 1092

```

QY            560   ACGGTAAGCTGATCGACACCATTCAAGTTCACCGGTATCAACTTCTCCAGA     619  
                |||        |||        |||        |||        |||        |||  
Db            1151   CGGGAGCGGTTCAACAGCCGAACAGAGTGTTGGACAGAACACGCCCGTGAAGA     1210  
  
QY            620   CCATCAGCTTCAGATCACAAAGATCCCAGACACCGTCTGATCAGCTCCGACTCCGACA     679

680 ACATCAACATGTGGATCCGTGACTTCTACATCTCCGCCAAGGAGTTGGACGGTAGGACA 739

[illegible]

```

QY 1040 ACGAGACCATGTACGCCGACACCACTCCACCGAGGCATCTAGCCCATCGG 1091
      ||      ||      ||      ||      ||      ||      ||      ||
Db 1631 GTGACTTCAACGGCGACGCAAGCGGCATCGGCGTCCTCTACGACTACGG 1682

RESULT 14
US-10-230-437-123

```

```

/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Gerritsen, Mary
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stephan, Jean-Philippe F.
/ APPLICANT: Watanabe, Colin L.
/ APPLICANT: Wood, William I.
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: ACIDS ENCODING THE SAME
/ FILE REFERENCE: P3530PIC94
/ CURRENT APPLICATION NUMBER: US/10/230,437
/ CURRENT FILING DATE: 2002-08-28
/ PRIOR APPLICATION NUMBER: 10/119,480
/ PRIOR FILING DATE: 2002-04-09
/ PRIOR APPLICATION NUMBER: 60/059113

```

; PRIOR FILING DATE: 1997-09-17  
 ; PRIOR APPLICATION NUMBER: 60/062287  
 ; PRIOR FILING DATE: 1997-10-17  
 ; PRIOR APPLICATION NUMBER: 60/063549

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

Db	1331	CGTTCACGGCAATGGCAACGGGTTTCAACAGCCCGGACAGCAGGTGTGGGACAGCAACAACG	1339
QY	800	ACCTGAGATACAACAAGAGTACTACATGSTCAACATCGACTACTTGAACAGATACATGT	859
Db	1391	ACCCCGTGAAGAGCTGGAACTGGGAAGCGCGCAAGCCCGTTCCGGCGGACTTCAACGGCG	1450
QY	860	AGCCCAACTCCAGACAGATCGTCTTCAACACAGACAGCTAAACAACAACGACTTCAACGAG	919
Db	1451	ACGGCAAGAGCGACATCGCGCTCTCTACGACATGGGTGCGCAACCGAGGACGGCGCAACC	1510
QY	920	GTACAGAGTATCATCAAGCGTATACAGAGGTATACCAACAACGACACCAAGTCTAGAGTG	979
Db	1511	GCACCAAGTGTTCACCTTCACCGGCACGGGACCGGTTTCAACAGCCCGGTCAAGTGT	1570
QY	980	GTGACATCTGTACTTCGACATGACTATCAACAACAGGCCTCAAACTGTTCTATGAAGA	1039
Db	1571	GGGACAGCAACGACGACCCCGGTGAAGAGCTGGAACGCTCGCGAGCAAGCCGCTCGCAG	1630
QY	1040	ACGAGACCATGTACCGCGACACCACTCCACCGAGGACATCTACGCCATCGG	1091
Db	1631	GTGACTTCAACGGCGGCAGCGGRCATCGGCTCTCTACGACTACGG	1682
RESULT 14			
US-10-230-437-123			
; Sequence 123, Application US/10230437			
; GENERAL INFORMATION:			
; APPLICANT: Baker, Kevin P.			
; APPLICANT: Desnoyers, Luc			
; APPLICANT: Gerritsen, Mary			
; APPLICANT: Goddard, Audrey			
; APPLICANT: Godowski, Paul J.			
; APPLICANT: Grimaldi, J. Christopher			
; APPLICANT: Gurney, Austin L.			
; APPLICANT: Smith, Victoria			
; APPLICANT: Stephan, Jean-Philippe F.			
; APPLICANT: Watanabe, Colin L.			
; APPLICANT: Wood, William T.			

```

; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 246
; SEQ ID NO 123
; LENGTH: 3401
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-230-437-123

Query Match
Best Local Similarity 44.3%; Score 44.4; DB 6; Length 3401;
Matches 180; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

730 GGTAAAGGACATCAACATCTGTTCACTCTTGCAGTACACCAAGTGTGTCAGGACTAC 789
983 GGGTGCACCTGCAGAGCTGTCCTCAACATGAGGGCACCAGCTCATGTCCTCAAC 1042
790 TGGGGTAAGCACTGAGATACAAAGAGAGTACTACATGGTCAACATCGACTACTTGAAC 849
1043 AGCCTCAAGAAGATGCGAAGCTGACTGAGTGGAGCTGATCCGCTGGAGCGC 1102
850 AGATACATGTACCCCACTCCAGACAGATCGTCTTCACACACAGAGCTTAACACACGAC 909
1103 ATCCCCCACTCCATCTTACGCTTCCCAACCTGCGAGAGATTGACCTCAAGGACAAAC 1162
910 TTCAAGGAGGGTTACAGATCATCATCAAGCGTATCAGAGGTAAACACCAAGCACCAGA 969
1163 CTCAAGACCATCGAGGAGATCATGACTTCCAGACCTGCACCGCTCACCTGCTTAAG 1222
970 GTGAGAGTGGTGACATCTGTTACTTGCAGATGACTATCAACAAAGGCTTACACCTG 1029
1223 CTGTGTACAAACCATCGCTTACATCCCATCCAGATCGGAGATTGACCTCAAGGACAG 1282
1030 TTCAAGAAACGAGACCATGTACGCCGACAAACCACTCCACCGAGGATCATGCCCATC 1089
1283 CGCCTCTACCTGAACCGCAACAGATCGAGAGATCCCAACCCAGCTCTTCTACTGCCG 1342
1090 GGCTGGTGAGACACCAAGGACATCAACGACACATCATCTTCC 1135
1343 AAGCTGGCTACCTGGAGCTCAGCCACAAACCTGACCTTCTCTCC 1388

RESULT 15
US-10-125-923A-411
Sequence 411, Application US/10125923A
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430RIC79
CURRENT APPLICATION NUMBER: US/10/125,923A
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250

```

```

; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 411
; LENGTH: 3401
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-125-923A-411

Query Match
Best Local Similarity 44.3%; Score 44.4; DB 6; Length 3401;
Matches 180; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 730 GGTAAAGGACATCAACATCTGTTCACTCTTGCAGTACACCAAGTGTGTCAGGACTAC 789
Db 983 GGGTGCACCTGCAGAGCTGTCCTCAACATGAGGGCACCAGCTCATGTCCTCAAC 1042
QY 790 TGGGGTAAGCACTGAGATACAAAGAGAGTACTACATGGTCAACATCGACTACTTGAAC 849
Db 1043 AGCCTCAAGAAGATGCGAAGCTGACTGAGTGGAGCTGATCCGCTGGAGCGC 1102
QY 850 AGATACATGTACCCCACTCCAGACAGATCGTCTTCACACACAGAGCTTAACACACGAC 909
Db 1103 ATCCCCCACTCCATCTTACGCTTCCCAACCTGCGAGAGATTGACCTCAAGGACAAAC 1162
QY 910 TTCAAGGAGGGTTACAGATCATCATCAAGCGTATCAGAGGTAAACACCAAGCACCAGA 969
Db 1163 CTCAAGACCATCGAGGAGATCATGACTTCCAGACCTGCACCGCTCACCTGCTTAAG 1222
QY 970 GTGAGAGTGGTGACATCTGTTACTTGCAGATGACTATCAACAAAGGCTTACACCTG 1029
Db 1223 CTGTGTACAAACCATCGCTTACATCCCATCCAGATCGGAGATTGACCTCAAGGACAG 1282
QY 1030 TTCAAGAAACGAGACCATGTACGCCGACAAACCACTCCACCGAGGATCATGCCCATC 1089
Db 1283 CGCCTCTACCTGAACCGCAACAGATCGAGAGATCCCAACCCAGCTCTTCTACTGCCG 1342
QY 1090 GGCTGGTGAGACACCAAGGACATCAACGACACATCATCTTCC 1135
Db 1343 AAGCTGGCTACCTGGAGCTCAGCCACAAACCTGACCTTCTCTCC 1388

Search completed: November 7, 2002, 14:42:56
Job time : 189 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:18:45 ; Search time 1949 Seconds  
(without alignments)  
11392.515 Million cell updates/sec

Title: US-09-910-186A-9

Perfect score: 1371

Sequence: 1 gaattacagatgaccatccc.....tctccgagtaatagaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154056 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estnu:\*

5: em\_estov:\*

6: em\_estopl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gas:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_man:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107.2	7.8	641	13	BM181884
2	107	7.8	906	17	BM153606
3	106.2	7.7	816	17	AZ533744
4	104.4	7.6	890	17	BH148886
5	104.4	7.6	908	17	AZ548467
6	104.2	7.6	905	17	AZ550256

c	7	100.6	7.3	976	17	BH149983	BH149983
c	8	100.4	7.3	843	17	AZ551618	ENTQD93TF
c	9	100.2	7.3	891	17	AZ683582	ENTK47TR
c	10	99.4	7.3	849	17	AZ546009	ENTFW53TF
c	11	98.8	7.2	494	13	BJ360881	BJ360881
c	12	98.8	7.2	681	17	CNS02E0D	ENTQD93TF
c	13	98.4	7.2	931	17	BH160272	ENTQD93TF
c	14	97.4	7.1	900	17	AZ549980	ENTQD93TF
c	15	96.8	7.1	797	13	BJ324597	BJ324597
c	16	95.8	7.0	608	13	BJ330059	BJ330059
c	17	95.4	7.0	877	17	AZ531291	ENTB034TR
c	18	95.4	7.0	912	17	AZ551092	ENTF122TF
c	19	92.6	6.8	356	13	BJ337970	BJ337970
c	20	92.6	6.8	645	9	AI389106	AI389106
c	21	92.4	6.7	603	13	BJ366946	BJ366946
c	22	92.2	6.7	500	17	B67199	B67199
c	23	91.8	6.7	450	13	BJ371297	BJ371297
c	24	91.4	6.7	711	13	BM588321	BM588321
c	25	90.6	6.6	551	13	BJ366220	BJ366220
c	26	90.2	6.6	317	13	BJ366180	BJ366180
c	27	90.2	6.6	394	9	AU060224	AU060224
c	28	89.8	6.5	511	13	BM004465	BM004465
c	29	89.8	6.5	827	17	CNS02156	CNS02156
c	30	89.4	6.5	649	13	BI641907	BI641907
c	31	89.2	6.5	527	10	AV962678	AV962678
c	32	88.8	6.5	880	17	AZ529191	AZ529191
c	33	86.2	6.3	644	13	BM605179	BM605179
c	34	85.6	6.2	585	13	BJ328723	BJ328723
c	35	85.2	6.2	888	17	AZ528430	AZ528430
c	36	84.8	6.2	789	17	AQ935224	AQ935224
c	37	84.8	6.2	890	17	AZ530768	AZ530768
c	38	84.2	6.1	542	13	BI594734	BI594734
c	39	83.8	6.1	494	13	BM641696	BM641696
c	40	83.6	6.1	359	13	BJ366856	BJ366856
c	41	83.6	6.1	503	13	BM656118	BM656118
c	42	83.4	6.1	878	17	BH153470	BH153470
c	43	83.4	6.1	970	17	CNS03H6V	CNS03H6V
c	44	82.6	6.0	607	13	BJ348812	BJ348812
c	45	81.8	6.0	608	13	BI887904	BI887904

#### ALIGNMENTS

BM181884 641 bp mRNA linear EST 11-DEC-2001  
fv51bl1.y1 Sugano SJD adult male Danio rerio cDNA clone 5412044 5'  
similar to contains element TARI repetitive element ;, mRNA  
sequence.

ACCESSION BM181884.1 GI:17512842  
VERSION BM181884  
KEYWORDS EST  
SOURCE zebrafish.  
ORGANISM zebrafish.

REFERENCE 1 (bases 1 to 641)  
AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, X., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; teleostei; Osteichthyes; Cypriniformes  
; Cyprinidae; Danio.

TITLE WashU Zebrafish EST Project 1998  
JOURNAL Unpublished (1998)  
COMMENT Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrafish@watson.wustl.edu





```

RESULT 4
BH146886/c
LOCUS
DEFINITION
  BH146886
  ENTFRK48TF Entamoeba histolytica Sheared DNA linear GSS 27-AUG-2001
  genomic, DNA sequence.
ACCESSION
  BH146886
  BH146886.1 GI:15302963
KEYWORDS
  GSS.
SOURCE
  Entamoeba histolytica.
  Entamoeba histolytica.
  Eukaryota; Entameobidae; Entamoeba.
REFERENCE
  1 (bases 1 to 890)
  Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  Determination of clone end sequences from Entamoeba histolytica
  HMI:IMSS sheared DNA library (2001)
  Unpublished (2001)
  CONTACT
  Contact: Brendan J Loftus
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0208
  Fax: 301 838 3543
  Email: b1ofustus@igrr.org
  Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
  DNA library
  Seq primer: M13-Forward
  Class: Shotgun
  High quality sequence start: 17
  High quality sequence stop: 880.
FEATURES
  Location/Qualifiers
    1..890
    /organism="Entamoeba histolytica"
    /strain="HMI:IMSS"
    /db_xref="taxon:5759"
    /clone_lib="Entamoeba histolytica Sheared DNA"
    /note="Vector: pHOSt; Site_1: 8st I; Constructed at The
    Institute for Genomic Research (TIGR), Rockville, MD.
    Genomic DNA isolated from broth cultures of E. histolytica
    using a method described by Clark and Diamond (Clark,
    C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
    method for isolate identification. Exp. Parasitol.
    77:450.). The DNA was mechanically sheared to give a
    tight size distribution (~2 kb). The v + i method used for
    the library construction is described in detail in Smith,
    H.O. and Venter, J.C. (Making small insert libraries for
    whole genome shotgun sequencing projects. In Genome
    Sequencing: A Practical Approach, eds. M. Vaubin and B.
    Barrel, Oxford University Press, 1999)."
```



```

/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHS1; Site1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      434 a   16 c   289 g   169 t
ORIGIN
Query Match      7.6%; Score 104.4; DB 17; Length 908;
Best Local Similarity 45.8%; Pred. No. 1.8e-12;
Matches 360; Conservative 0; Mismatches 426; Indels 0; Gaps 0;
4 TTACGATGACCAATCCATTCACATCTCTCTACCAACAACATCCCTGTGAAGGAC 63
819 TTTCATCATCTTCATCATCTTCATCATCTCTCTCTCTCTCTCTCTCTCTCT 760
64 ATCATCAACGAGTACTTCACACATCAACAGCTCCCAAGATCTCTCTCTCTCT 123
759 TTCTTCATCATCTTCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 700
124 AAGAACCTTGTGTCGACCTCCGTTTACACGGCGAGGTCTCCGAGGAGGTGAC 183
699 ATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 640
184 CAGCTGAACCAATCTCCCATCTTCAGCTCTCTCTCTCTCTCTCTCTCTCT 243
639 GTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 580
244 AAGTCTATCTGTCACCCAGAGACATCTCTCTCTCTCTCTCTCTCTCTCT 303
579 ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 520
304 ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 363
519 ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 460
364 GACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 423
459 TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 400
424 ACCCTGAGCAGACGAGGACTCGGAGCAGTCCATCACTCTCTCTCTCTCT 483
399 TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 340
484 AACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 543
339 ATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 280
544 ATGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 603
279 TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 220
604 ATCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 663
219 ATCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 160
664 ACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 723
159 GTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 100
724 TTGACGGTGAAGGACATCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 783
99 TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 40

```

```

QY 784 GACTAC 789
Db 39 ATCTTC 34

RESULT 6
AZ550256
LOCUS ENT5E58TR Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000
DEFINITION genomic, DNA sequence.
ACCESSION AZ550256
VERSION AZ550256.1 GI:11175557
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 905)
AUTHORS Loftus B., Van Aken S. and Fraser C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI-IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI-IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 100
High quality sequence stop: 872.
Location/Qualifiers
1..905
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHS1; Site1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT      203 a   260 c   46 g   396 t
ORIGIN
Query Match      7.6%; Score 104.2; DB 17; Length 905;
Best Local Similarity 48.2%; Pred. No. 2e-12;
Matches 326; Conservative 0; Mismatches 348; Indels 3; Gaps 1;
QY 270 CATCGTCTACAAGTCCATTCAGAGTCTCTCTCTCTCTCTCTCTCTCT 329
Db 150 CGTCATCTTCATCAACATCTTCGTCATCTTCATCATCTTCATCATCAT 209
QY 330 GTGGGTCTCAACTGCCAGGTACACCATCATCGACTCGGTCAAGAACTCG 389
Db 210 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 269
QY 390 GTCCATCGGTATCATCCCAACTTCGTGGTCTTCACCTCGAAGCAAGAGACT 449
Db 270 CTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTTCATCTCT 329

```

Qy 450 CCACTGATCAACTTCTCCCTACGACATCTCCACACAGGCTCGCTGGTTACAAAGAGTGT 509  
 Db 330 CGTCTTCATCATCATCTTCATCTTCATCTTCTTAATCAAAATTCATCATCTTCTT 389  
 Qy 510 CTTGGTCAACGCTCACCACACATGATGGGTAAACATGAGATATACATCAACGGTAAGCT 569  
 Db 390 CTTCTTCATCTTCATCATCATCAT--CACTTCATCACTTTCATCATCTTCATCAT 446  
 Qy 570 GATGACACATCAAGGTCAAGAGTGAACGGTATCACTTCTCCAAACCATCACTT 629  
 Db 447 CTTTCATCTTCATCATCTTCATCATCATCATCTTCTTCATCTTCTTCATCTT 506  
 Qy 630 CGAGATCAACAGATCCAGACACGCTGATCACTCGACTCCGACGACATCAACAT 689  
 Db 507 CTTTCATCTTCTTAATTCATCATCATCTTCTTCTTCTTCTTCTTCATCATCAT 566  
 Qy 690 GTGATCCGAGTACTACATCTTCGCAAGAGTGGAGGTAAAGCATCAACATCTT 749  
 Db 567 CATCTTCTTCATCTTCATCATCATCTTTCATCATCATCTTCTTCATCATCATCTT 626  
 Qy 750 GTTCAACTCTTCGAGTACACACAGGTCGTAAGGACTACTGGGTAAAGCTGAGATA 809  
 Db 627 CATCTTCTTCATCTTCTTAATTCATCATCATCTTCTTCTTCTTCTTCTTCATCAT 586  
 Qy 810 CAACAAGAGTACTACATCTTCGATCTACTTGAACAGATACATGATGACCAACTC 869  
 Db 687 CATCATCTTCTTCTTCATCTTCTTCGCTTCATCATCATCATCTTCTTCTTCTTCT 746  
 Qy 870 CAGACAGATCTTTCACACACAGGCTTACACACAGGCTTCAACGAGGTTTACAGAT 929  
 Db 747 ATTCAATTCATCATCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 806  
 Qy 930 CATCATCAAGCGTATCA 946  
 Db 807 CCAATTCATCACTTCTCA 823

RESULT 7  
 LOCUS BH149983 976 bp DNA linear GSS 27-AUG-2001  
 DEFINITION ENTQD93TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 genomic, DNA sequence.  
 ACCESSION BH149983  
 VERSION BH149983.1 GI:15311165  
 WORDS GSS.  
 ORGANISM Entamoeba histolytica.  
 Entamoeba histolytica  
 Eukaryota; Entamoebidae; Entamoeba.  
 1 (bases 1 to 976)  
 Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.  
 Determination of clone end sequences from Entamoeba histolytica  
 HMI:IMSS sheared DNA library (2001)  
 Unpublished (2001)  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: b.loftus@igrr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
 DNA library  
 Seq primer: M13-Forward  
 Class: shotgun  
 High quality sequence start: 38  
 High quality sequence stop: 740.  
 Location/Qualifiers  
 1. 976  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHS1; Site\_1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaubin and B.  
 Barrel, Oxford University Press, 1999))."

BASE COUNT 415 a 63 c 280 g 218 t  
 ORIGIN  
 Query Match 7.3%; Score 100.6; DB 17; Length 976;  
 Best Local Similarity 48.1%; Pred. No. 1.3e-11;  
 Matches 317; Conservative 0; Mismatches 339; Indels 3; Gaps 1;  
 Qy 247 GTCATCGTCCACGACGAGACATCGTCTCAACTCCATCCATGATGAGTCTTCCTCCATC 306  
 Db 786 GTCCTTCATCGTCTTCGATCGTCATCTCGATCGTCTTCCATCGTCTTCTTCATC 727  
 Qy 307 TCCCTTCGGATCAGATCAACAAGTGGTCTCCAACTTCCAGGTACACCATCATCGAC 366  
 Db 726 GTCCTTCATCGTCTTCATCATCTTCTTCATCATCATCTTCATCATCATCTTCATCAT 667  
 Qy 367 TCCGTCAAGAACAACTCCGGTTCGTCATCGGTATCATCTCCAACTTCCCTGCTCTTACC 426  
 Db 666 ATCTTCATCATCTTCTTCATCATCATCTTCTTCATCATCATCTTCTTCATCATCATC 607  
 Qy 427 CTGAAGCAGACGAGGACTCCGAGCAGTCACTTCTTCATCATCATCTTCATCATCATCT 486  
 Db 606 TTCCTTCATCATCTTCATCTTCAT---CATCATCTTCTTCATCATCATCTTCATCATC 550  
 Qy 487 GTCCTGTTACAAACAGTGGTCTTCGTCACCGTCCACAAACATGATGGTAAACATG 546  
 Db 549 GTCCTTCAATTCATCACTTTCATCATCATCTTCATCATCATCTTCTTCATCATCT 490  
 Qy 547 AAGATCAGATCAACGGTAGCTGATCGACACATCAAGTCAAGGAGTTCACCGGTATC 506  
 Db 489 ATCGTCTTCATCATCATCTTCATCTTCATCATCTTCATCATCTTCATCATCTTCATC 430  
 Qy 607 AACTTCTCCAGACCATCACCTTCGAGATCAACAAGATCCACACACCGGTCTGTATCAC 666  
 Db 429 ATCGTCTTCATCATCATCTTCATCTTCATCATCTTCATCATCTTCATCATCTTCATC 370  
 Qy 667 TCCGACTCCGACACATCAACATGATGGATCCGAGTCTTCATCTTCATCTTCGCAAGAGTTG 726  
 Db 369 ATCGTCTTCATCATCATCTTCATCTTCATCATCTTCATCATCTTCATCTTCATCTTCAT 310  
 Qy 727 GACGGTAAGGACATCAACATCTTCCTTCACTCTTCAGTACACCAACGTCGTCAGGAC 786  
 Db 309 TCTTCATCTTCCTTCCTTCATCTTCCTTCATCTTCCTTCATCTTCCTTCATCATCTC 250  
 Qy 787 TACGGGGTAACGACCTGAGATCAACAAGAGTACTTACATGCTCAACATCGACTACTTGG 846  
 Db 249 TTTCTGCTGCTTCTTAATTCAAATTCATCATCTTCATCTTCATCTTCATCTTCATCTTAA 190  
 Qy 847 AACAGATACATGACGCCCAACTCCAGACAGATCTTCCTTCAACACAGACGATCAACAA 905  
 Db 189 TTCATCTCATCATCATTAATAAATTCATTAATTCATTAATTCATTAATTCATCATCTCGA 131  
 RESULT 8  
 LOCUS AZ551618 843 bp DNA linear GSS 14-NOV-2000  
 DEFINITION ENTQD54TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
 genomic, DNA sequence.  
 ACCESSION AZ551618  
 VERSION AZ551618.1 GI:11176919  
 WORDS GSS.  
 SOURCE Entamoeba histolytica.

663	QY	CACCTCGGACTCGGACAAACATGATGGATCCGTCGACTTCATCACTTCGCCAAGGA	722
664	QY		
665	QY		
666	QY		
667	QY		
668	QY		
669	QY		
670	QY		
671	QY		
672	QY		
673	QY		
674	QY		
675	QY		
676	QY		
677	QY		
678	QY		
679	QY		
680	QY		
681	QY		
682	QY		
683	QY		
684	QY		
685	QY		
686	QY		
687	QY		
688	QY		
689	QY		
690	QY		
691	QY		
692	QY		
693	QY		
694	QY		
695	QY		
696	QY		
697	QY		
698	QY		
699	QY		
700	QY		
701	QY		
702	QY		
703	QY		
704	QY		
705	QY		
706	QY		
707	QY		
708	QY		
709	QY		
710	QY		
711	QY		
712	QY		
713	QY		
714	QY		
715	QY		
716	QY		
717	QY		
718	QY		
719	QY		
720	QY		
721	QY		
722	QY		
723	QY		
724	QY		
725	QY		
726	QY		
727	QY		
728	QY		
729	QY		
730	QY		
731	QY		
732	QY		
733	QY		
734	QY		
735	QY		
736	QY		
737	QY		
738	QY		
739	QY		
740	QY		
741	QY		
742	QY		
743	QY		
744	QY		
745	QY		
746	QY		
747	QY		
748	QY		
749	QY		
750	QY		
751	QY		
752	QY		
753	QY		
754	QY		
755	QY		
756	QY		
757	QY		
758	QY		
759	QY		
760	QY		
761	QY		
762	QY		
763	QY		
764	QY		
765	QY		
766	QY		
767	QY		
768	QY		
769	QY		
770	QY		
771	QY		
772	QY		
773	QY		
774	QY		
775	QY		
776	QY		
777	QY		
778	QY		
779	QY		
780	QY		
781	QY		
782	QY		
783	QY		
784	QY		
785	QY		
786	QY		
787	QY		
788	QY		
789	QY		</

BASE COUNT	199 a	236 c	43 g	371 t	ORIGIN
Query Match	7.3%	Score 99.4;	DB 17;	Length 849;	
Best Local Similarity	47.7%;	Pred. NO. 2.2e-11;			
Matches 326;	Conservative 0;	Mismatches 351;	Indels 6;	Gaps 1;	
QY	270	CATCGCTACAACTCCCATGTCTCTTCCATCTCTCTTCTGGATCAGAAATCAACAA	329		
Db	49	CGTCATCTTCATCAACATCTTCGTCATCTTCATCACTTCATCATCATCATCAT	108		
QY	330	GTGGGTCTCCAACTWTGGAGGTTACACCATCATCTCGACTCGGTCCAGAACAACTCGGGTTG	389		
Db	109	CTTCTTTCTCATCTTCATCATCATCTTTCTTCTTCTTCTTCTTCATCATCTTCATCTTCTT	168		
QY	390	GTCCATCGGTATCATCTCCCAACTTCCCGGTCTTCACCCCTGAACAGACAGAGACTCCGA	449		
Db	169	CTTCATCTTCATCTTCATCTTTCATCTTAATTCAAATTCATCATCATCTTCGTCTTCTT	228		
QY	450	GGAGTCCATACAACTTCTCCTTACGACATCTCCACACAGCGTCCCTGGTTACACAACTGGTT	509		
Db	229	CGTCTTCATCATCATCTTCATCTTCTTCTTCATCTTCTTAATTCAAATTCATCATCATCTTCTT	288		

Db	289	CITCTTCATCTTCATCATCATCATCATCTTCTCACTCATCTTTCATCATCTT	348
QY	564	TAAGTGTATCGACACCATCAAGGTCAAGGTATGACCACTTCTCCAAGACAT	623
Db	349	CATCATCTTCATCATCTTTCATCATCTTTCATCATCATCATCATCTTCTT	408
QY	624	CACCTTCGAGATCAACAGATCCACAGACCGGTCTCATCACTCGCATCCGACAAT	683
Db	409	CATCTTCTTCATCTTCTTAATTCAAATTCATCATCATCATCTTCTTCTGCTTCTT	468
QY	684	CAACATGTGGATCCGTGACTTCTACATCTTTCGCCAAGGAGTTGGACGGTAAGACAT	743

```
Db 459 CATCATCATCTCTTCATCTTCATCATCATCTTCATCATCATCATCTCTCTTCTCATCATCAT 528
Oy 744 CATCTGTTCACCTCTTGGAGTACACCAACGCTGCTCAAGGACTACTGGGGTAACGACCT 803
Db 529 CATCTTCATCTCTTCATCTCTTCAATCAAAATCATCATCATCTCTCTCTCTCTCTCTCT 588
Oy 804 GAGATACACAGAGAGTACTACATGTTCAACATCGACTACTTGAACAGATACATGTACGC 863
Db 589 CATCATCATCATCTCTCTCTCTTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 648
Oy 864 CAATCTCCAGACAGATCTCTCTCAACACGACGCTTCAACAGACTTCAACGAGGGTTA 923
Db 649 CTCTTAATCAATTCATCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 708
Oy 924 CAAGATCATCATCAAGGATCA 946
Db 709 CATGTTCCATTTTCATATTTTCA 731

RESULT 11
LOCUS BJ360881
DEFINITION BJ360881 Dictyostelium discoidium cDNA library, CF Dictyostelium
ACCESSION BJ360881
VERSION BJ360881.1 GI:19260476
KEYWORDS EST.
SOURCE Dictyostelium discoidium.
ORGANISM Dictyostelium discoidium.
REFERENCE 1 (bases 1 to 494)
AUTHORS Eukaryote; Mycetozoa; Dictyostelidia; Dictyostelium.
TITLES Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
FULL LENGTH cDNA of Dictyostelium discoidium at the culmination
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
LOCATION/Qualifiers
FEATURES
source
1. 494
/organism="Dictyostelium discoidium"
/strain="AX4"
/db_xref="taxon:44689"
/clone="gdc8k12"
/sex="mat A"
/dev_stage="Culmination stage"
BASE COUNT 279 a 129 c 34 g 48 t 4 others
ORIGIN
Query Match 7.2%; Score 98.8; DB 13; Length 494;
Best Local Similarity 52.2%; Pred. No. 2.4e-11;
Matches 214; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

Oy 680 ACATCAACATGTGGATCGTGTACATCTTCGCGCAAGGAGTGGCGGTAAAGACA 739
Db 80 ACATGAAGATATAGATTGGTTTAAATTTCTTGATAGTTTACATTCATGTTGTAACGA 139
Oy 740 TCAACATCTGTTCACATCCCTTGCAGTACACCAACGCTCGTCAAGGACTACTGGGTAACG 799
Db 140 ACATATTCACACACACACACACACACACACACACACACACACACACACACACACACAC 199
Oy 800 ACCTGAGTACACACACACACACACACACACACACACACACACACACACACACACACAC 859
Db 200 ACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 259
Oy 860 AGGCAACTCCAGACAGATCTCTCTCAACACACACACACACACACACACACACACACAC 919
Db 260 GCAGCAACACACACACACACACACACACACACACACACACACACACACACACACACAC 319
```

```
Oy 920 GTTACAGATCATCATCAACGCGTATCAGAGGTAAACCAACGACACACACAGTCAAGGTG 979
Db 320 GCAACAAGACACACACACACACACACACACACACACACACACACACACACACACACAC 379
Oy 980 GTGACATCTCTGTCTCGACATGACTATCAACCAAGGCTTACAACTGTTTCATGAAGA 1039
Db 380 ACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 439
Oy 1040 ACAGACCATGTAGCGCGACACACACCTCCACCGAGGACATCTACGCCATC 1089
Db 440 ACAAACAACACACACACACACACACACACACACACACACACACACACACACACACAC 489

RESULT 12
CNS02EOD 681 bp DNA linear GSS 13-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 262H14 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL193990
VERSION AL193990.1 GI:7832096
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 681)
AUTHORS Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brotter, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLES Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 681)
AUTHORS Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
TITLES Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 681)
AUTHORS Direct Submission
TITLES Submitted (12-APR-2000)
JOURNAL This sequence is a single read and was generated as part of a large
COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1. 681
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="262H14"
/clone_lib="G"
/notes="Genoscope sequence ID : C0AG262DD07LPI-end : T7"
BASE COUNT 74 a 105 c 170 g 282 t 50 others
ORIGIN
Query Match 7.2%; Score 98.8; DB 17; Length 681;
Best Local Similarity 48.8%; Pred. No. 2.7e-11;
Matches 197; Conservative 20; Mismatches 187; Indels 0; Gaps 0;

Oy 813 CAAGAGTACTACATGTTCAACATCTGACTTCAACAGATACATGTACGCCAATCCAG 872
Db 406 CDACACACACACACACACACACACACACACACACACACACACACACACACACACAC 347
Oy 873 ACAGATCTCTTCAACACACACACACACACACACACACACACACACACACACACACAC 932
Db 346 CAACGACACACACACACACACACACACACACACACACACACACACACACACACACAC 287
```

```

Qy 933 CATCAACGGTATCAGAGGTAAACACCAACGACACCAAGTACAGAGGTGGTGACATCCTGTA 992
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 CAGCGAGGAGGAGCAGCAGCAACAAACAAACAAACAAACAAACAAACAAACAAACAA 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 993 CTTCCGATGACTATCAACAAAGGCGCTACACCTGTTTCATGAAGAGGAGACCATGTA 1052
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 CAACAACAACAACAACAACAACACACACACACACACACACACACACACACACACAA 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1053 CGCGCAACACACTCCACCGAGGACATCTACGCCATCGGTCTCGGTGAGCAGCAAGGA 1112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 CRACAACAACACACACACACACACACACACACACACACACACACACACACACAA 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1113 CATCAAGGACACATCATCTCCAGATCCAGCCAAAGTACACACATCTACTAGCTC 1172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 AAACAACAACAACACACACACACACACACACACACACACACACACACACAA 47
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1173 CCAGATCTTCAAGTCCAACTTCAACGGTGAAGACATCTCGGTA 1216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
46 CAACAACAACAACACACACACACACACACACACACACACACACACACACAACTTKA 3

RESULT 13
LOCUS BH160272/c
DEFINITION ENTQ49TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION BH160272
VERSION BH160272.1 GI:15733710
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 931)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 8
High quality sequence stop: 677.
Location/Qualifiers
1..931
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHO51; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

```

Best Local Similarity 47.0%; Pred. No. 3.8e-11;
Matches 303; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

Qy 249 CATCGTCACCCAGAACGAGACATCGTCTACAACTCCATGTACGAGTCTCTTCTCCATCTC 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 562 CGTCGCTTCATCATCTTTCATCATCTTTCATCATCTTTCATCATCTTTCATCATCTTCT 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 309 CTTCTGATCAGAAATCAACAAGTGGGTCTCCAACTTCCAGGTTCCAGGTTACCAACATCAGCTC 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 602 CTTTCATCATCTTCTTTCATCATCTTTCATCATCTTTCATCATCTTTCATCATCTTCT 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 369 CGTCAAGAACAATCCGGTGGTCCATCGGTATCATCTCCAACTTCCCTGGTCTTCCACCT 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 542 CATCTTCTTCTTTCATCATCTTCTTTCATCATCTTCTTTCATCATCTTCTTTCATCAT 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 429 GAGCAGAACGAGACTCGGAGAGTCCAGTCAACTTCTCTAGGACATCTCCCAACAGC 488
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 CTTCTAATTCATATTCGTCTTCATCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 489 TCCGGTTTACAACAAGTGGTCTTCTCGTCCCGTCCACCAACATGATGGGTAAACATGAA 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 CATCTTCTTCTTTCATCATCTTCTTTCATCATCTTCTTTCATCATCATCTTCTCAT 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 549 GATCTACATCAACGGTAAGTGTATGACGACACCACTCAAGGTCAAGGAGTTCACGGTATCAA 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 CTTCTTCTTTCATCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 609 CTTCTCCAAAGCATTCCCTCGAGATCAACAGATCCAGACACCGGTCTGATCACCTC 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 CTTCTTCTTTCATCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 669 CGACTCCGACAACTCAACATGTGGTGGTACTTCTATCTTCTTCTTCTTCTTCTTCTTCT 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 CATCGTCTTTCATCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 729 CGGTAAGGACATCAACATCTGTTCAACTCTTTCAGTACACCAAGCTCGTCAAGACTA 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 CTTCTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 789 CTGGGTGAAGACCTGAGATCAACAAGGAGTACTACATGGTCAACACTGACTACTTGA 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 CATCATCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 849 CAGATCATGTAGCCCAACCCAGACAGATCGTCTTCAACACA 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 CGTCTTCATCATCTTTCATCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 19
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
LOCUS AZ549980/c
DEFINITION ENTDD94TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION AZ549980
VERSION AZ549980.1 GI:11175122
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 900)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
```

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 20

High quality sequence stop: 890.

## FEATURES

source

1. .900

/organism="Entamoeba histolytica"

```
/strain="HMI:IMSS"
```

```
/db xref="taxon:5759"
```

```
/usr_xref = taxon:3733
/clone lib="Entamoeba histolytica sheared DNA"
```

Genomic DNA isolated from broth cultures of *E. histolytica* using a method described by Clark and Diamond (Clark, C.C., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + 1 method used for

10

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barell, Oxford University Press, 1999)."

63 c	235 q	226 t
------	-------	-------

2

7.18; Score 97.4; DB 17; Length 900;

7.19; score 57.4; DB 17; remych 500;  
ty 49.08; Pred. No. 6.2e-11;

Accuracy	45.0%	Freq. NO: 0.25	IT
Derivative	0	Mismatches	271
		Indels	0
		Gaps	0

[illegible]

TAACGAGAAACATCGTCTACAACCTCCATGTACGAGTCCTTCTCCATCTCCT 310

[illegible]

GTCTTCATCATCGTCTTCATCATCGTCTTCATCATCGTCTTCATCATCG 801

GGCTTCATCATCGGCTTTCATCATCGGCTTTCATCATCG

ATCAACAAGTGGGTCTCCAACCTGCCAGGTTACACCATCATCGACTCCG 370

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

GTCTTCATCATCGTCTTCATCATCGTCTTCATCATCG 741

CGCTTCATCAGTGGCTTCATCAGTGCTTCATCAGG / 41

TTCCGGGTGGTCCATCCGATATCATCTCCAACTTCCCTGGTCTTCACCCCTGA 430

TCGGGTGGTCCATCGGTATCATCTCCAACTTCCCTGGTCTTCACCTTGA 430

GTGTCATCATCGTCCTCATTCGTCTCATTCGTCTCATTCG 681

05/15/2017

43T

431 AGCAGAACGAGGAGCTCCGAGCAGTCCATCAACTTCTCTACGAGATCTCCAAACAAGGCTC 430

[illegible]

**089**

197

491 C TGGT TACAACAAGTGGTCTCTCGTCAACCGTCAACACATGATGGTAAACA TGAAGA 550

620 ECE 561

